UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/959,935	11/13/2001	Simon J Powers	35-1511	6850

7590 07/22/2004

Nixon & Vanderhye
1100 North Glebe Road 8th Floor
Arlington, VA 22201-4714

EXAMINER	
BROSS, EDWARD J	

ART UNIT	PAPER NUMBER
2126	

DATE MAILED: 07/22/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/959,935	POWERS ET AL.	
	Examiner	Art Unit	
	Edward Bross	2126	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *13 November 2001*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-10* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-10* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☒ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☒ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

1. Claims 1-10 are pending in this application.

2. Claims 4, 5, 9 and 10 are objected to under 37 CFR 1.75(c) as being in improper form because a multiple dependent claim cannot depend from any other multiple dependent claim. See MPEP § 608.01(n). Accordingly, the claims have not been further treated on the merits.

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

4. Claims 1, 2, 6 and 7 are rejected under 35 U.S.C. 103(a) as being unpatentable over Tafoya et al. (6,411,988).

5. As to claims 1 and 6, Tafoya discloses a method of providing communication between two or more software elements, a host computer means being arranged to host application programs (00 and 50 Fig. 2A) to host application software elements in a host space, two or more application software elements being hosted in said host space (10, 22, 23, 22 and 23 Fig. 2A); the method comprising:

associating each application software element with a communication software element through which to send and/or receive messages (21 Fig. 2A, col. 5 lines 44-47);

allowing each application software element to communicate with other application software elements by sending and receiving messages through the respectively associated communication software elements (col. 5 lines 44-47);

allowing each software application software element and associated software element to move in said host space (i.e. the implicit moving of such elements in memory during swapping and other memory management performed by the operating system).

6. Tafoya does not explicitly disclose holding the communication state of each associated application software element in its associated communication software element. However, storing the communication state in a communication software element is well known in the art (i.e. the state of the underlying socket as stored in a Socket object in a Java application).

7. It would have been obvious to one of ordinary skill in the art at the time of the invention to store the communication state in the associated communication software elements of Tafoya as this would have increased the modularity of the component thus increasing the maintainability of the code.

8. As to claims 2 and 7, Tafoya does not explicitly discloses the step of holding the communication state of the associated application software element comprises holding a queue

of messages not yet delivered to the associated application software element. However, the use of message queues to hold undelivered messages is well known in the art.

9. It would have been obvious to one of ordinary skill in the art at the time of the invention to use a message queue to hold messages not yet delivered to the associated software element in the system of Tafoya in order to increase the performance by allowing for asynchronous delivery of messages between software elements.

10. Claims 3 and 8 are rejected under 35 U.S.C. 103(a) as being unpatentable over Tafoya et al. (6,411,988) in view of Jagannathan (6,496,871).

11. As to claims 3 and 8, Tafoya discloses host computer means comprises two or more host computers (col. 5, lines 35-39).

12. Tafoya does not disclose the step of allowing each application software element and associated communication software element to move in said host space comprises allowing each application software element and associated communication software element to move between said two or more host computers.

13. Jagannathan discloses application software elements moving between two or more host computers (abstract).

14. It would have been obvious to one of ordinary skill in the art at the time of the invention to use the software element relocation of Jagannathan with the system of Tafoya in order to

increase the performance of the system by allowing the migration of the elements to a les loaded computer.

15. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Edward Bross whose telephone number is 703-305-8754. The examiner can normally be reached on Mon-Fri 8:30-5:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Meng-Ai An can be reached on 703-305-9678. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

EB

Notice of References Cited	Application/Control No. 09/959,935	Applicant(s)/Patent Under Reexamination POWERS ET AL.	
	Examiner Edward Bross	Art Unit 2126	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-6,411,988	06-2002	Tafoya et al.	709/204
	B	US-6,496,871	12-2002	Jagannathan et al.	719/317
	C	US-			
	D	US-			
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

BIBDATASHEET

CONFIRMATION NO. 6850

Bib Data Sheet

SERIAL NUMBER	FILING DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
09/959,935	11/13/2001 (RULE	709	2126	35-1511

APPLICANTS

Simon J Powers, Ipswich, Suffolk, GBN, UNITED KINGDOM;

Michael R Hinds, Felixstowe, Suffolk, GBN, UNITED KINGDOM;

** CONTINUING DATA **********

This application is a 371 of PCT/GB00/02245 06/09/2000

OK GB

** FOREIGN APPLICATIONS **********

EUROPEAN PATENT OFFICE (EPO) 99304559.0 06/11/1999

OK GB

Foreign Priority claimed ☒ yes ☐ no 35 USC 119 (a-d) conditions met ☒ yes ☐ no ☐ Met after Allowance Verified and Acknowledged Examiner's Signature Initials	STATE OR COUNTRY GBN	SHEETS DRAWING 8	TOTAL CLAIMS 10	INDEPENDENT CLAIMS 2

ADDRESS

Nixon & Vanderhye
1100 North Glebe Road 8th Floor
Arlington , VA
22201-4714

TITLE

Communication between software elements

FILING FEE RECEIVED 890	FEES: Authority has been given in Paper No. _____ to charge/credit DEPOSIT ACCOUNT No. _____ for following:	☐ All Fees ☐ 1.16 Fees (Filing) ☐ 1.17 Fees (Processing Ext. of time) ☐ 1.18 Fees (Issue) ☐ Other _____ ☐ Credit

Index of Claims

Application No.	Applicant(s)	
09/959,935	POWERS ET AL.	
Examiner	**Art Unit**	
Edward Bross	2126	

√	Rejected	–	(Through numeral) Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

Claim		Date								
Final	Original	7/9/04								
	1	√								
	2	√								
	3	√								
	4	√								
	5	√								
	6	√								
	7	√								
	8	√								
	9	√								
	10	√								
	11									
	12									
	13									
	14									
	15									
	16									
	17									
	18									
	19									
	20									
	21									
	22									
	23									
	24									
	25									
	26									
	27									
	28									
	29									
	30									
	31									
	32									
	33									
	34									
	35									
	36									
	37									
	38									
	39									
	40									
	41									
	42									
	43									
	44									
	45									
	46									
	47									
	48									
	49									
	50									

Claim		Date								
Final	Original									
	51									
	52									
	53									
	54									
	55									
	56									
	57									
	58									
	59									
	60									
	61									
	62									
	63									
	64									
	65									
	66									
	67									
	68									
	69									
	70									
	71									
	72									
	73									
	74									
	75									
	76									
	77									
	78									
	79									
	80									
	81									
	82									
	83									
	84									
	85									
	86									
	87									
	88									
	89									
	90									
	91									
	92									
	93									
	94									
	95									
	96									
	97									
	98									
	99									
	100									

Claim		Date								
Final	Original									
	101									
	102									
	103									
	104									
	105									
	106									
	107									
	108									
	109									
	110									
	111									
	112									
	113									
	114									
	115									
	116									
	117									
	118									
	119									
	120									
	121									
	122									
	123									
	124									
	125									
	126									
	127									
	128									
	129									
	130									
	131									
	132									
	133									
	134									
	135									
	136									
	137									
	138									
	139									
	140									
	141									
	142									
	143									
	144									
	145									
	146									
	147									
	148									
	149									
	150									

Search Notes

Application No.	Applicant(s)	
09/959,935	POWERS ET AL.	
Examiner	**Art Unit**	
Edward Bross	2126	

SEARCHED			
Class	Subclass	Date	Examiner
709	204, 203, 205, 317	7/9/2004	EB

INTERFERENCE SEARCHED			
Class	Subclass	Date	Examiner

SEARCH NOTES (INCLUDING SEARCH STRATEGY)		
	DATE	EXMR
EAST, ACM, IEEE	7/9/2004	EB

AU Zumstein E; Pearson B M; Kalogeropoulos A; Schweizer M
CS Institute of Food Research, Genetics & Microbiology Department, Norwich Research Park, Colney, U.K.
SO Yeast (Chichester, England), **(1995 Aug)** 11 (10) 975-86.
Journal code: 8607637. ISSN: 0749-503X.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
OS GENBANK-M73270; GENBANK-X83121
EM 199601
ED Entered STN: 19960220
Last Updated on STN: 19960220
Entered Medline: 19960126
AB The nucleotide sequence of a 29.425 kb fragment localized on the left arm of chromosome XV from Saccharomyces cerevisiae has been determined. The sequence contains 13 open reading frames (ORFs) of which four encode the known genes ADH1, COQ3, MSH2 and RCF4. Predictions are made concerning the functions of the unknown ORFs. Some of the ORFs contain sequences similar to expressed sequence tags (EST) found in the **database** made available by **TIGR**. In particular, the highly expressed ADH1 gene is represented in this **database** by no less than 20 EST sequences. Two ARS sequences and a putative functional GCN4 motif have also been detected. One ORF (O0953) containing nine putative transmembrane segments is similar to a hypothetical membrane protein of Arabidopsis thaliana. Characteristic features of the other ORFs include ATP/GTP binding sites, a fungal Zn(2)-Cys(6) binuclear centre, an endoplasmic reticulum targeting sequence, a beta-transducin repeat signature and in two instances, good similarity to the prokaryotic lipoprotein signal peptide motif.

```
Connecting via Winsock to STN



Welcome to STN International!  Enter x:x

LOGINID:SSSPTA1805JXB

PASSWORD:
TERMINAL (ENTER 1, 2, 3, OR ?):2


* * * * * * * * * *   Welcome to STN International   * * * * * * * * * *

 NEWS  1          Web Page URLs for STN Seminar Schedule - N. America
 NEWS  2          "Ask CAS" for self-help around the clock
 NEWS  3  May 12  EXTEND option available in structure searching
 NEWS  4  May 12  Polymer links for the POLYLINK command completed in REGISTRY
 NEWS  5  May 27  New UPM (Update Code Maximum) field for more efficient patent
                  SDIs in CAplus
 NEWS  6  May 27  CAplus super roles and document types searchable in REGISTRY
 NEWS  7  Jun 22  STN Patent Forums to be held July 19-22, 2004
 NEWS  8  Jun 28  Additional enzyme-catalyzed reactions added to CASREACT
 NEWS  9  Jun 28  ANTE, AQUALINE, BIOENG, CIVILENG, ENVIROENG, MECHENG,
                  and WATER from CSA now available on STN(R)
 NEWS 10  Jul 12  BEILSTEIN enhanced with new display and select options,
                  resulting in a closer connection to BABS

 NEWS EXPRESS  MARCH 31 CURRENT WINDOWS VERSION IS V7.00A, CURRENT
               MACINTOSH VERSION IS V6.0c(ENG) AND V6.0Jc(JP),
               AND CURRENT DISCOVER FILE IS DATED 26 APRIL 2004
 NEWS HOURS    STN Operating Hours Plus Help Desk Availability
 NEWS INTER    General Internet Information
 NEWS LOGIN    Welcome Banner and News Items
 NEWS PHONE    Direct Dial and Telecommunication Network Access to STN
 NEWS WWW      CAS World Wide Web Site (general information)

Enter NEWS followed by the item number or name to see news on that
specific topic.

  All use of STN is subject to the provisions of the STN Customer
  agreement.  Please note that this agreement limits use to scientific
  research.  Use for software development or design or implementation
  of commercial gateways or other similar uses is prohibited and may
  result in loss of user privileges and other penalties.

  * * * * * * * * * * * * * * STN  Columbus  * * * * * * * * * * * * * * *

FILE 'HOME' ENTERED AT 09:56:59 ON 20 JUL 2004

=> file .pub
COST IN U.S. DOLLARS                                SINCE FILE      TOTAL
                                                       ENTRY    SESSION
FULL ESTIMATED COST                                     0.21       0.21

FILE 'MEDLINE' ENTERED AT 09:57:22 ON 20 JUL 2004

FILE 'BIOSIS' ENTERED AT 09:57:22 ON 20 JUL 2004
COPYRIGHT (C) 2004 BIOLOGICAL ABSTRACTS INC.(R)
```

```
=> s sequence and database and repeat
L1        1868 SEQUENCE AND DATABASE AND REPEAT

=> s l1 and (iterative or mask?)
L2          34 L1 AND (ITERATIVE OR MASK?)

=> s sequence and database and repeat?
L3        2174 SEQUENCE AND DATABASE AND REPEAT?

=> s l3 and (iterative or mask?)
L4          44 L3 AND (ITERATIVE OR MASK?)

=> duplicate remove l4
DUPLICATE PREFERENCE IS 'MEDLINE, BIOSIS'
KEEP DUPLICATES FROM MORE THAN ONE FILE?  Y/(N):n
PROCESSING COMPLETED FOR L4
L5           28 DUPLICATE REMOVE L4 (16 DUPLICATES REMOVED)

=> d 1-28 bib ab

L5   ANSWER 1 OF 28     MEDLINE on STN                    DUPLICATE 1
AN   2004099923     MEDLINE
DN   PubMed ID: 14990456
TI   AntiHunter: searching BLAST output for EST antisense transcripts.
AU   Lavorgna Giovanni; Sessa Luca; Guffanti Alessandro; Lassandro Lelio;
     Casari Giorgio
CS   Istituto Scientifico H. S. Raffaele, Via Olgettina 60, 20132 Milan,
     Italy.. giovanni.lavorgna@hsr.it
SO   Bioinformatics (Oxford, England), (2004 Mar 1) 20 (4) 583-5.
     Journal code: 9808944. ISSN: 1367-4803.
CY   England: United Kingdom
DT   (EVALUATION STUDIES)
     Journal; Article; (JOURNAL ARTICLE)
     (VALIDATION STUDIES)
LA   English
FS   Priority Journals
EM   200406
ED   Entered STN: 20040302
     Last Updated on STN: 20040625
     Entered Medline: 20040624
AB   AntiHunter is a new web-based tool for the identification of expressed
     sequence tag (EST) antisense transcripts from BLAST output.  In
     order to perform an analysis, user is required to input a genomic
     sequence plus an associated list of transcript names and
     coordinates of the genomic region (i.e. genome annotation).  After
     masking the repeated regions (if any), program will
     perform a BLASTN search of the input sequence versus the
     selected EST database, reporting by Email the EST entries that
     reveal a putative antisense transcript with respect to the user supplied
     list.

L5   ANSWER 2 OF 28  BIOSIS  COPYRIGHT 2004 BIOLOGICAL ABSTRACTS INC. on STN
AN   2004:174674  BIOSIS
DN   PREV200400176053
TI   Transposable element annotation of the rice genome.
AU   Juretic, Nikoleta [Reprint Author]; Bureau, Thomas E.; Bruskiewich,
     Richard M.
CS   Department of Biology, McGill University, Montreal, PQ, H3A 1B1, Canada
     njuret@po-box.mcgill.ca
SO   Bioinformatics (Oxford), (January 22 2004) Vol. 20, No. 2, pp. 155-160.
     print.
     ISSN: 1367-4803.
```

DT Article
LA English
ED Entered STN: 31 Mar 2004
Last Updated on STN: 31 Mar 2004
AB Motivation: The high content of repetitive **sequences** in the genomes of many higher eukaryotes renders the task of annotating them computationally intensive. Presently, the only widely accepted method of searching and annotating transposable elements (TEs) in large genomic **sequences** is the use of the RopeatMasker program, which identifies new copies of TEs by pairwise **sequence** comparisons with a library of known TEs. Profile hidden Markov models (HMMs) have been used successfully in discovering distant homologs of known proteins in large protein **databases**, but this approach has only rarely been applied to known model TE families in genomic DNA. Results: We used a combination of computational approaches to annotate the TEs in the finished genome of Oryza sativa ssp. japonica. In this paper, we discuss the strengths and the weaknesses of the annotation methods used. These approaches included: the default configuration of **Repeat-Masker** using cross-match, an implementation of the Smith-Waterman-Gotoh algorithm; RopeatMasker using WU-BLAST for similarity searching; and the HMMER package, used to search for TEs with profile HMMs. All the results were converted into GFF format and post-processed using a set of Perl scripts. **RepeatMasker** was used in the case of most TE families. The WU-BLAST implementation of **RepeatMasker** was found to be manifold faster than cross-match with only a slight loss in sensitivity and was thus used to obtain the final set of data. HMMER was used in the annotation of the Mutator-like element (MULE) superfamily and the miniature inverted-**repeat** transposable element (MITE) polyphyletic group of families, for which large libraries of elements were available and which could be divided into well-defined families. The HMMER search algorithm was extremely slow for models over 1000 bp in length, so MULE families with members over 1000 bp long were processed with **RepeatMasker** instead. The main disadvantage of HMMER in this application is that, since it was developed with protein **sequences** in mind, it does not search the negative DNA strand. With the exception of TE families with essentially palindromic **sequences**, reverse complement models had to be created and run to compensate for this shortcoming. We conclude that a modification of **RepeatMasker** to incorporate libraries of profile HMMs in searches could improve the ability to detect degenerated copies of TEs.

L5 ANSWER 3 OF 28 MEDLINE on STN DUPLICATE 2
AN 2003297210 MEDLINE
DN PubMed ID: 12824401
TI ESTAnnotator: A tool for high throughput EST annotation.
AU Hotz-Wagenblatt Agnes; Hankeln Thomas; Ernst Peter; Glatting Karl-Heinz; Schmidt Erwin R; Suhai Sandor
CS Department of Molecular Biophysics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany.. hotz-wagenblatt@dkfz.de
SO Nucleic acids research, (2003 Jul 1) 31 (13) 3716-9.
Journal code: 0411011. ISSN: 1362-4962.
CY England: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200308
ED Entered STN: 20030626
Last Updated on STN: 20030819
Entered Medline: 20030818
AB In high throughput **sequence** analysis, it is often necessary to combine the results of contemporary bioinformatics tools, because no

individual tool alone computes all the requested information. ESTAnnotator is a tool for the high throughput annotation of expressed **sequence** tags (ESTs) by automatically running a collection of bioinformatics applications. In the first step, a quality check is performed and **repeats**, vector parts and low quality **sequences** are **masked**. Then successive steps of **database** searching and EST clustering are performed. Already known transcripts present within mRNA and genomic DNA reference **databases** are identified. Subsequently, tools for the clustering of anonymous ESTs, and for further **database** searches at the protein level, are applied. Finally, the outputs of each individual tool are gathered and the relevant results presented in a descriptive summary. ESTAnnotator was already successfully applied for the systematic identification and characterisation of novel human genes involved in cartilage/bone formation, growth, differentiation and homeostasis. ESTAnnotator is available at http://genome.dkfz-heidelberg.de, contact: genome@dkfz.de.

L5 ANSWER 4 OF 28 MEDLINE on STN
AN 2003476960 MEDLINE
DN PubMed ID: 14555627
TI Modes and clustering for time-warped gene expression profile data.
AU Liu Xueli; Muller Hans-Georg
CS Department of Human Genetics, UCLA School of Medicine, Los Angeles, CA 90095, USA.
SO Bioinformatics (Oxford, England), (2003 Oct 12) 19 (15) 1937-44. Journal code: 9808944. ISSN: 1367-4803.
CY England: United Kingdom
DT (EVALUATION STUDIES)
Journal; Article; (JOURNAL ARTICLE)
(VALIDATION STUDIES)
LA English
FS Priority Journals
EM 200406
ED Entered STN: 20031015
Last Updated on STN: 20040630
Entered Medline: 20040629
AB MOTIVATION: The study of the dynamics of regulatory processes has led to increased interest for the analysis of temporal gene expression level data. To address the dynamics of regulation, expression data are collected **repeatedly** over time. It is difficult to statistically represent the resulting high-dimensional data. When regulatory processes determine gene expression, time-warping is likely to be present, i.e. the sample of gene expression trajectories reflects variation not only in terms of the expression amplitudes, but also in terms of the temporal structure of gene expression. RESULTS: A non-parametric time-synchronized **iterative** mean updating technique is proposed to find an overall representation that corresponds to a mode of a sample of expression profiles, viewed as a random sample in function space. The proposed algorithm explores the application of previous work of Hall and Heckman to genome-wide expression data and provides an extension that includes random time-warping with the aim to synchronize timescales across genes. The proposed algorithm is universally applicable for the construction of modes for functional data with time-warping. We demonstrate the construction of mode functions for a sample of Drosophila gene expression data. The algorithm can be applied to define clusters among the observed trajectories of gene expression, without any kind of prior non-time-warped clustering, as illustrated in the numerical example.

L5 ANSWER 5 OF 28 BIOSIS COPYRIGHT 2004 BIOLOGICAL ABSTRACTS INC. on STN
AN 2004:140161 BIOSIS

DN PREV200400133613

TI The significance of published polymorphisms in seven cases of mild factor VII deficiency.

AU Cutler, Jacqueline A. [Reprint Author]; Patel, Rinku [Reprint Author]; Mitchell, Michael J. [Reprint Author]; Rangarajan, Savita [Reprint Author]; Savidge, Geoffrey F. [Reprint Author]

CS Haemophilia Centre, Reference Centre for Haemostatic and Thrombotic Disorders, St Thomas' Hospital, London, UK

SO Blood, (November 16 2003) Vol. 102, No. 11, pp. 308a. print.
Meeting Info.: 45th Annual Meeting of the American Society of Hematology. San Diego, CA, USA. December 06-09, 2003. American Society of Hematology.
CODEN: BLOOAW. ISSN: 0006-4971.

DT Conference; (Meeting)
Conference; (Meeting Poster)
Conference; Abstract; (Meeting Abstract)

LA English

ED Entered STN: 10 Mar 2004
Last Updated on STN: 10 Mar 2004

AB FVII deficiency is the most common of the 'rare inherited coagulation disorders', having an estimated prevalence of 1:400,000. It is an autosomal recessive disorder, characterised by epistaxes, gum bleeding and menorraghia. Patients with severe FVII deficiency may suffer joint bleeds and, more commonly, bleeding into the central nervous system. There is only a weak correlation between coagulant activity and clinical bleeding tendency, and patients with very low levels of plasma FVII may exhibit fewer symptoms than others with much higher levels. FVII is involved in the initiation of the coagulation cascade, as its activation, by complexing with exposed tissue factor leads to the activation of factors IX and X. FVII levels are determined by both environmental and genetic factors, and a number of polymorphisms have been identified which may be associated with a reduced level of FVII. Arg353Gln has a 0.2% allele frequency in Caucasians. Heterozygosity for this polymorphism results in a 20-25% reduction in both FVII antigen and activity through impaired secretion from hepatocytes. A decanucleotide deletion/insertion at position -323 of the promotor (-323 0/10), is found in linkage with Arg353Gln in some populations, and has an allele frequency of 0.23%. Study of a polish population, in which the two polymorphisms do not express strong linkage, has established that these polymorphisms are independently associated with a reduction in factor FVII levels, and that their effects are approximately equal, but are not additive. The deletion/insertion polymorphism affects the rate of transcription, and so may **mask** the effects of reduced secretion. A variable number of **repeats** of a 37 base pair **sequence** within intron 7 has also been linked with variation in plasma FVII levels. The rare IVS7+7A-G polymorphism is located within the first **repeat**, adjacent to the IVS7 donor splice site. Although originally believed to be functionally silent, it is now thought to modify splicing by an as yet undetermined mechanism. Analysis of patients with identical numbers of IVS7 **repeats** showed that the lowest FVII levels were in those patients who had the IVS7+7G allele. Over a two-year period twelve patients were referred to this centre for molecular analysis following detection of low factor VII levels. Of these, a molecular defect was identified in only five. Four of these patients had missense mutations, and FVII levels reduced to 37-45% of normal, whilst the fifth patient was heterozygous for two missense mutations with resultant FVII level of 14% of normal. In the remaining seven patients, all of whom presented with bleeding symptoms and reduced factor VII levels (58-81% of normal), no 'molecular abnormality' was identified following sequencing of the entire coding and promotor regions of the factor VII gene. Re-evaluation of these patients showed that all possessed one or more of the polymorphisms described above. Our findings suggest that the presence of one or more of these polymorphisms is the genetic basis for the reduced FVII:C in each of these patients.

The significance of these polymorphisms has been understated in the Factor VII **database**, where, due to relatively high allele frequencies, they are listed as polymorphisms despite published evidence of their effect on FVII levels. We would recommend that the presence of these polymorphisms is investigated first in all cases of mild FVII deficiency.

L5 ANSWER 6 OF 28 MEDLINE on STN DUPLICATE 3
AN 2003283170 MEDLINE
DN PubMed ID: 12809672
TI An exhaustive DNA micro-satellite map of the human genome using high performance computing.
AU Collins Jack R; Stephens Robert M; Gold Bert; Long Bill; Dean Michael; Burt Stanley K
CS Advanced Biomedical Computing Center, NCI-Frederick, Frederick, MD, USA.
NC N01-C0-12400
SO Genomics, (2003 Jul) 82 (1) 10-9.
Journal code: 8800135. ISSN: 0888-7543.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200402
ED Entered STN: 20030618
Last Updated on STN: 20040224
Entered Medline: 20040223
AB The current pace of the generation of **sequence** data requires the development of software tools that can rapidly provide full annotation of the data. We have developed a new method for rapid **sequence** comparison using the exact match algorithm without **repeat masking**. As a demonstration, we have identified all perfect simple tandem **repeats** (STR) within the draft **sequence** of the human genome. The STR elements (chromosome, position, length and **repeat** subunit) have been placed into a relational **database**. **Repeat** flanking **sequence** is also publicly accessible at http://grid.abcc.ncifcrf.gov. To illustrate the utility of this complete set of STR elements, we documented the increased density of potentially polymorphic markers throughout the genome. The new STR markers may be useful in disease association studies because so many STR elements manifest multiallelic polymorphism. Also, because triplet **repeat** expansions are important for human disease etiology, we identified trinucleotide **repeats** that exist within exons of known genes. This resulted in a list that includes all 14 genes known to undergo polynucleotide expansion, and 48 additional candidates. Several of these are non-polyglutamine triplet **repeats**. Other examinations of the STR **database** demonstrated **repeats** spanning splice junctions and identified SNPs within **repeat** elements.

L5 ANSWER 7 OF 28 BIOSIS COPYRIGHT 2004 BIOLOGICAL ABSTRACTS INC. on STN
AN 2003:582716 BIOSIS
DN PREV200300572542
TI IN SILICO GENE IDENTIFICATION IN COLONIC NEOPLASIA.
AU Moss, Alan [Reprint Author]; Madden, Stephen [Reprint Author]; Mathuna, Padraic Mac [Reprint Author]; Doran, Peter [Reprint Author]
CS Dublin, Ireland
SO Digestive Disease Week Abstracts and Itinerary Planner, (2003) Vol. 2003, pp. Abstract No. 812. e-file.
Meeting Info.: Digestive Disease 2003. FL, Orlando, USA. May 17-22, 2003. American Association for the Study of Liver Diseases; American Gastroenterological Association; American Society for Gastrointestinal Endoscopy; Society for Surgery of the Alimentary Tract.
DT Conference; (Meeting)

Conference; Abstract; (Meeting Abstract)
LA English
ED Entered STN: 10 Dec 2003
Last Updated on STN: 10 Dec 2003
AB BACKGROUND: Digital Differential Display (DDD) is a computational strategy for the identification of cDNAs whose expression is altered in different tissue types or pathological states. This technology exploits the vast amount of cDNA libraries available in public **sequence databases**. Comparisons of these libraries allow enriched cDNAs to be identified. A major limitation of this approach, in common with many **database**-mining strategies, is the poor annotation of the **sequence** libraries. Here we describe the application of a high-throughput annotation pipeline to analysis of colonic neoplasia libraries.AIMS: In this study we have employed an integrated bioinformatics-based approach to (a) identify genes whose expression is altered in colon cancer libraries (b) annotate cDNAs without homology to known genes that are identified as disease-associated. METHODS: EST libraries from normal and neoplastic colon were compared using Digital Differential Display, resulting in the compilation of gene lists that are exclusively expressed, statistically significant, or preferentially expressed in colon cancer. Transcripts without homology to known genes were annotated using a novel platform, Digital Extractor, which was developed in-house. This solution integrates and utilizes a number of tools including; (a) CAP3, for assembly of EST clusters, (b) **RepeatMasker** to **mask** repetitive elements and (c) BLAST, for gene identification.RESULTS: DDD comparison of colon cancer libraries to normal colon and normal adult tissue identified 204 ESTs altered in colon cancer. 38 of these genes have previously been described in colon cancer (for example APOBEC1, GPA33). 127 represent known genes that have not previously been identified in colon cancer (for example ETV4, TRIM31). Furthermore, 39 cDNAs without homology to known genes were identified. Annotation of these data resulted in the identification of several known genes (for example CDX2, Ribosomal protein L41).CONCLUSION: This novel computational biology-based approach can identify genes differentially expressed in colon cancer. The novel proteins are currently being validated ex vivo..

L5 ANSWER 8 OF 28 MEDLINE on STN DUPLICATE 4
AN 2002609982 MEDLINE
DN PubMed ID: 12364612
TI A comparison of profile hidden Markov model procedures for remote homology detection.
AU Madera Martin; Gough Julian
CS MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.. mm238@mrc-lmb.cam.ac.uk
SO Nucleic acids research, (2002 Oct 1) 30 (19) 4321-8.
Journal code: 0411011. ISSN: 1362-4962.
CY England: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200212
ED Entered STN: 20021008
Last Updated on STN: 20021218
Entered Medline: 20021216
AB Profile hidden Markov models (HMMs) are amongst the most successful procedures for detecting remote homology between proteins. There are two popular profile HMM programs, HMMER and SAM. Little is known about their performance relative to each other and to the recently improved version of PSI-BLAST. Here we compare the two programs to each other and to non-HMM methods, to determine their relative performance and the features that are important for their success. The quality of the multiple **sequence**

alignments used to build models was the most important factor affecting the overall performance of profile HMMs. The SAM T99 procedure is needed to produce high quality alignments automatically, and the lack of an equivalent component in HMMER makes it less complete as a package. Using the default options and parameters as would be expected of an inexpert user, it was found that from identical alignments SAM consistently produces better models than HMMER and that the relative performance of the model-scoring components varies. On average, HMMER was found to be between one and three times faster than SAM when searching **databases** larger than 2000 **sequences**, SAM being faster on smaller ones. Both methods were shown to have effective low complexity and **repeat sequence masking** using their null models, and the accuracy of their E-values was comparable. It was found that the SAM T99 **iterative database** search procedure performs better than the most recent version of PSI-BLAST, but that scoring of PSI-BLAST profiles is more than 30 times faster than scoring of SAM models.

L5 ANSWER 9 OF 28 MEDLINE on STN
AN 2002631254 MEDLINE
DN PubMed ID: 12389629
TI AINT/ERIC/TACC: an expanding family of proteins with C-terminal coiled coil domains.
AU Lappin Terence R; Mullan Robert N; Stewart J Peter; Morgan Neal A; Thompson Alexander; Maxwell A Peter
CS Haematology Group, Cancer Research Centre, Queen's University Belfast, Belfast City Hospital, UK.. t.lappin@qub.ac.uk
SO Leukemia & lymphoma, (2002 Jul) 43 (7) 1455-9. Ref: 16
Journal code: 9007422. ISSN: 1042-8194.
CY England: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
General Review; (REVIEW)
(REVIEW, TUTORIAL)
LA English
FS Priority Journals
EM 200309
ED Entered STN: 20021023
Last Updated on STN: 20030903
Entered Medline: 20030902
AB The AINT/ERIC/TACC genes encode novel proteins with a coiled coil domain at their C-terminus. The founding member of this expanding family of genes, transforming acidic coiled coil 1 (TACC1), was isolated from a BAC contig spanning the breast cancer amplicon-1 on 8p11. Transfection of cells in vitro with TACC1 resulted in anchorage-independent growth consistent with a more "neoplastic" phenotype. **Database** searches employing the human TACC1 **sequence** revealed other novel genes, TACC2 and TACC3, with substantial **sequence** homology particularly in the C-terminal regions encoding the coiled coil domains. TACC2, located at 10q26, is similar to anti-zuai-1 (AZU-1), a candidate breast tumour suppressor gene, and ECTACC, an endothelial cell TACC which is upregulated by erythropoietin (Epo). The murine homologue of TACC3, murine erythropoietin-induced cDNA (mERIC-1) was also found to be upregulated by Epo in the Friend virus anaemia (FVA) model by differential display-PCR. Human ERIC-1, located at 4p16.3, has been cloned and encodes an 838-amino acid protein whose N- and C-terminal regions are highly homologous to the shorter 558-amino acid murine protein, mERIC-1. In contrast, the central portions of these proteins differ markedly. The murine protein contains four 24 amino acid imperfect **repeats**. ARNT interacting protein (AINT), a protein expressed during embryonic development in the mouse, binds through its coiled coil region to the aryl hydrocarbon nuclear translocator protein (ARNT) and has a central portion that contains seven of the 24 amino acid **repeats** found in

mERIC-1. Thus mERIC-1 and AINT appear to be developmentally regulated alternative transcripts of the gene. Most members of the TACC family discovered so far contain a novel nine amino acid putative phosphorylation site with the pattern [R/K]-X(3)-[E]-X(3)-Y. Genes with **sequence** homology to the AINT/ERIC/TACC family in other species include **maskin** in Xenopus, D-TACC in Drosophila and TACC4 in the rabbit. **Maskin** contains a peptide **sequence** conserved among eIF-4E binding proteins that is involved in oocyte development. D-TACC cooperates with another conserved microtubule-associated protein Msps to stabilise spindle poles during cell division. The diversity of function already attributed to this protein family, including both transforming and tumour suppressor properties, should ensure that a new and interesting narrative is about to unfold.

L5 ANSWER 10 OF 28 MEDLINE on STN DUPLICATE 5
AN 2002657639 MEDLINE
DN PubMed ID: 12417195
TI Digging deep for ancient relics: a survey of protein motifs in the intergenic **sequences** of four eukaryotic genomes.
AU Zhang Zhao Lei; Harrison Paul M; Gerstein Mark
CS Department of Molecular Biophysics and Biochemistry, Yale University, Bass Center 432A, 266 Whitney Avenue, P.O. Box 208114, New Haven, CT 06520-8114, USA.
SO Journal of molecular biology, (2002 Nov 8) 323 (5) 811-22.
Journal code: 2985088R. ISSN: 0022-2836.
CY England: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200212
ED Entered STN: 20021106
Last Updated on STN: 20021218
Entered Medline: 20021213
AB We have examined conserved protein motifs in the non-coding, intergenic regions ("pseudomotif patterns") and surveyed their occurrence in the fly, worm, yeast and human genomes (chromosomes 21 and 22 only). To identify these patterns, we **masked** out annotated genes, pseudogenes and **repeat** regions from the raw genomic **sequence** and then compared the remaining **sequence**, in six-frame translation, against 1319 patterns from the PROSITE **database**. For each pseudomotif pattern, the absolute number of occurrences is not very informative unless compared against a statistical expectation; consequently, we calculated the expected occurrence of each pattern using a Poisson model and verified this with simulations. Using a p-value cut-off of 0.01, we found 67 pseudomotif patterns over-represented in fly intergenic regions, 34 in worm, 21 in human and six in yeast. These include the zinc finger, leucine zipper, nucleotide-binding motif and EGF domain. Many of the over-represented patterns were common to two or more organisms, but there were a few that were unique to specific ones. Furthermore, we found more over-represented patterns in the fly than in the worm, although the fly has fewer pseudogenes. This puzzling observation can be explained by a higher deletion rate in the fly genome. We also surveyed under-represented patterns, finding 23 in the fly, 12 in the worm, 18 in human and two in yeast. If intergenic **sequences** were truly random, we would expect an equal number of over and under-represented patterns. The fact that for each organism the number of over-represented patterns is greater than the number of under-represented ones implies that a fraction of the intergenic regions consist of ancient protein fragments that, due to accumulated disablements, have become unrecognizable by conventional techniques for gene and pseudogene identification. Moreover, we find that in aggregate the over-represented pseudomotif patterns occupy a substantial fraction of the intergenic

regions. Further information is available at http://pseudogene.org

L5 ANSWER 11 OF 28 MEDLINE on STN
AN 2001246740 MEDLINE
DN PubMed ID: 11238992
TI Gene2EST: a BLAST2 server for searching expressed **sequence** tag (EST) **databases** with eukaryotic gene-sized queries.
AU Gemund C; Ramu C; Altenberg-Greulich B; Gibson T J
CS European Molecular Biology Laboratory, Postfach 10.2209, 69012 Heidelberg, Germany.
SO Nucleic acids research, (2001 Mar 15) 29 (6) 1272-7.
Journal code: 0411011. ISSN: 1362-4962.
CY England: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200105
ED Entered STN: 20010517
Last Updated on STN: 20010521
Entered Medline: 20010510
AB Expressed **sequence** tags (ESTs) are randomly sequenced cDNA clones. Currently, nearly 3 million human and 2 million mouse ESTs provide valuable resources that enable researchers to investigate the products of gene expression. The EST **databases** have proven to be useful tools for detecting homologous genes, for exon mapping, revealing differential splicing, etc. With the increasing availability of large amounts of poorly characterised eukaryotic (notably human) genomic **sequence**, ESTs have now become a vital tool for gene identification, sometimes yielding the only unambiguous evidence for the existence of a gene expression product. However, BLAST-based Web servers available to the general user have not kept pace with these developments and do not provide appropriate tools for querying EST **databases** with large highly spliced genes, often spanning 50 000-100 000 bases or more. Here we describe Gene2EST (http://woody.embl-heidelberg.de/gene2est/), a server that brings together a set of tools enabling efficient retrieval of ESTs matching large DNA queries and their subsequent analysis. **RepeatMasker** is used to **mask** dispersed repetitive **sequences** (such as Alu elements) in the query, BLAST2 for searching EST **databases** and Artemis for graphical display of the findings. Gene2EST combines these components into a Web resource targeted at the researcher who wishes to study one or a few genes to a high level of detail.

L5 ANSWER 12 OF 28 BIOSIS COPYRIGHT 2004 BIOLOGICAL ABSTRACTS INC. on STN
AN 2001:192824 BIOSIS
DN PREV200100192824
TI Gene2EST: A BLAST2 server for searching expressed **sequence** tag (EST) **databases** with eukaryotic gene-sized queries.
AU Gemuend, Christine; Ramu, Chenna; Altenberg-Greulich, Brigitte; Gibson, Toby J. [Reprint author]
CS European Molecular Biology Laboratory, 69012, Heidelberg, Germany
toby.gibson@embl-heidelberg.de
SO Nucleic Acids Research, (March 15, 2001) Vol. 29, No. 6, pp. 1272-1277. print.
CODEN: NARHAD. ISSN: 0305-1048.
DT Article
LA English
ED Entered STN: 20 Apr 2001
Last Updated on STN: 18 Feb 2002
AB Expressed **sequence** tags (ESTs) are randomly sequenced cDNA clones. Currently, nearly 3 million human and 2 million mouse ESTs provide valuable resources that enable researchers to investigate the

products of gene expression. The EST **databases** have proven to be useful tools for detecting homologous genes, for exon mapping, revealing differential splicing, etc. With the increasing availability of large amounts of poorly characterised eukaryotic (notably human) genomic **sequence**, ESTs have now become a vital tool for gene identification, sometimes yielding the only unambiguous evidence for the existence of a gene expression product. However, BLAST-based Web servers available to the general user have not kept pace with these developments and do not provide appropriate tools for querying EST **databases** with large highly spliced genes, often spanning 50 000-100 000 bases or more. Here we describe Gene2EST (http://woody.embl-heidelberg.de/gene2est/), a server that brings together a set of tools enabling efficient retrieval of ESTs matching large DNA queries and their subsequent analysis. **Repeat-Masker** is used to **mask** dispersed repetitive **sequences** (such as Alu elements) in the query, BLAST2 for searching EST **databases** and Artemis for graphical display of the findings. Gene2EST combines these components into a Web resource targeted at the researcher who wishes to study one or a few genes to a high level of detail.

L5 ANSWER 13 OF 28 MEDLINE on STN DUPLICATE 6
AN 2002116055 MEDLINE
DN PubMed ID: 11846551
TI Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily.
AU Wrabl J O; Grishin N V
CS Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas TX 75390-9050, USA.
SO Journal of molecular biology, (2001 Nov 30) 314 (3) 365-74.
Journal code: 2985088R. ISSN: 0022-2836.
CY England: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200202
ED Entered STN: 20020220
Last Updated on STN: 20030321
Entered Medline: 20020227
AB The O-linked GlcNAc transferases (OGTs) are a recently characterized group of largely eukaryotic enzymes that add a single beta-N-acetylglucosamine moiety to specific serine or threonine hydroxyls. In humans, this process may be part of a sugar regulation mechanism or cellular signaling pathway that is involved in many important diseases, such as diabetes, cancer, and neurodegeneration. However, no structural information about the human OGT exists, except for the identification of tetratricopeptide **repeats** (TPR) at the N terminus. The locations of substrate binding sites are unknown and the structural basis for this enzyme's function is not clear. Here, remote homology is reported between the OGTs and a large group of diverse sugar processing enzymes, including proteins with known structure such as glycogen phosphorylase, UDP-GlcNAc 2-epimerase, and the glycosyl transferase MurG. This relationship, in conjunction with amino acid similarity spanning the entire length of the **sequence**, implies that the fold of the human OGT consists of two Rossmann-like domains C-terminal to the TPR region. A conserved motif in the second Rossmann domain points to the UDP-GlcNAc donor binding site. This conclusion is supported by a combination of statistically significant PSI-BLAST hits, consensus secondary structure predictions, and a fold recognition hit to MurG. Additionally, **iterative** PSI-BLAST **database** searches reveal that proteins homologous to the OGTs form a large and diverse superfamily that is termed GPGTF (glycogen phosphorylase/glycosyl transferase). Up to one-third of the 51 functional families in the CAZY **database**, a glycosyl transferase classification scheme based on

catalytic residue and **sequence** homology considerations, can be unified through this common predicted fold. GPGTF homologs constitute a substantial fraction of known proteins: 0.4% of all non-redundant **sequences** and about 1% of proteins in the Escherichia coli genome are found to belong to the GPGTF superfamily.

L5 ANSWER 14 OF 28 MEDLINE on STN DUPLICATE 7
AN 2000493598 MEDLINE
DN PubMed ID: 10958633
TI Alfresco--a workbench for comparative genomic **sequence** analysis.
AU Jareborg N; Durbin R
CS The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom.. niclas.jareborg@cgr.ki.se
SO Genome research, (2000 Aug) 10 (8) 1148-57.
Journal code: 9518021. ISSN: 1088-9051.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200010
ED Entered STN: 20001027
Last Updated on STN: 20001027
Entered Medline: 20001017
AB Comparative analysis of genomic **sequences** provides a powerful tool for identifying regions of potential biologic function; by comparing corresponding regions of genomes from suitable species, protein coding or regulatory regions can be identified by their homology. This requires the use of several specific types of computational analysis tools. Many programs exist for these types of analysis; not many exist for overall view/control of the results, which is necessary for large-scale genomic **sequence** analysis. Using Java, we have developed a new visualization tool that allows effective comparative genome **sequence** analysis. The program handles a pair of **sequences** from putatively homologous regions in different species. Results from various different existing external analysis programs, such as **database** searching, gene prediction, **repeat masking**, and alignment programs, are visualized and used to find corresponding functional **sequence** domains in the two **sequences**. The user interacts with the program through a graphic display of the genome regions, in which an independently scrollable and zoomable symbolic representation of the **sequences** is shown. As an example, the analysis of two unannotated orthologous genomic **sequences** from human and mouse containing parts of the UTY locus is presented.

L5 ANSWER 15 OF 28 MEDLINE on STN
AN 2001164239 MEDLINE
DN PubMed ID: 11159316
TI **MaskerAid**: a performance enhancement to **RepeatMasker**.
AU Bedell J A; Korf I; Gish W
CS Genome Sequencing Center and Department of Genetics, Washington University School of Medicine, St. Louis, MO 63108, USA.
NC P50HG01458 (NHGRI)
SO Bioinformatics (Oxford, England), (2000 Nov) 16 (11) 1040-1.
Journal code: 9808944. ISSN: 1367-4803.
CY England: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200104
ED Entered STN: 20010425

Last Updated on STN: 20010425
Entered Medline: 20010419

AB SUMMARY: Identifying and **masking** repetitive elements is usually the first step when analyzing vertebrate genomic **sequence**. Current **repeat** identification software is sensitive but slow, creating a costly bottleneck in large-scale analyses. We have developed **MaskerAid**, a software enhancement to **RepeatMasker** that increased the speed of **masking** more than 30-fold at the most sensitive setting. AVAILABILITY: On request from the authors (see http://sapiens.wustl.edu/**MaskerAid**). CONTACT: **maskeraid**@watson.wustl.edu

L5 ANSWER 16 OF 28 MEDLINE on STN DUPLICATE 8
AN 2001092141 MEDLINE
DN PubMed ID: 11023840
TI Proteins of the endoplasmic-reticulum-associated degradation pathway: domain detection and function prediction.
AU Ponting C P
CS MRC Functional Genetics Unit, Department of Human Anatomy and Genetics, University of Oxford, South Parks Road, Oxford OX1 3QX, UK.. Chris.Ponting@anat.ox.ac.uk
SO Biochemical journal, (2000 Oct 15) 351 Pt 2 527-35.
Journal code: 2984726R. ISSN: 0264-6021.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200101
ED Entered STN: 20010322
Last Updated on STN: 20030215
Entered Medline: 20010125

AB **Sequence database** searches, using **iterative**-profile and Hidden-Markov-model approaches, were used to detect hitherto-undetected homologues of proteins that regulate the endoplasmic reticulum (ER)-associated degradation pathway. The translocon-associated subunit Sec63p (Sec=secretory) was shown to contain a domain of unknown function found twice in several Brr2p-like RNA helicases (Brr2=bad response to refrigeration 2). Additionally, Cue1p (Cue=coupling of ubiquitin conjugation to ER degradation), a yeast protein that recruits the ubiquitin-conjugating (UBC) enzyme Ubc7p to an ER-associated complex, was found to be one of a large family of putative scaffolding-domain-containing proteins that include the autocrine motility factor receptor and fungal Vps9p (Vps=vacuolar protein sorting). Two other yeast translocon-associated molecules, Sec72p and Hrd3p (Hrd=3-hydroxy-3-methylglutaryl-CoA reductase degradation), were shown to contain multiple tetratricopeptide-**repeat**-like **sequences**. From this observation it is suggested that Sec72p associates with a heat-shock protein, Hsp70, in a manner analogous to that known for Hop (Hsp70/Hsp90 organizing protein). Finally, the luminal portion of Ire1p (Ire=high inositol-requiring), thought to convey the sensing function of this transmembrane kinase and endoribonuclease, was shown to contain **repeats** similar to those in beta-propeller proteins. This finding hints at the mechanism by which Ire1p may sense extended unfolded proteins at the expense of compact folded molecules.

L5 ANSWER 17 OF 28 MEDLINE on STN DUPLICATE 9
AN 2000237731 MEDLINE
DN PubMed ID: 10772867
TI Homology-based method for identification of protein **repeats** using statistical significance estimates.
AU Andrade M A; Ponting C P; Gibson T J; Bork P
CS European Molecular Biology Laboratory, Meyerhofstr. 1, Heidelberg, 69012,

Germany.
SO Journal of molecular biology, (2000 May 5) 298 (3) 521-37.
Journal code: 2985088R. ISSN: 0022-2836.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals; Space Life Sciences
EM 200005
ED Entered STN: 20000613
Last Updated on STN: 20000613
Entered Medline: 20000530
AB Short protein **repeats**, frequently with a length between 20 and 40 residues, represent a significant fraction of known proteins. Many **repeats** appear to possess high amino acid substitution rates and thus recognition of **repeat** homologues is highly problematic. Even if the presence of a certain **repeat** family is known, the exact locations and the number of repetitive units often cannot be determined using current methods. We have devised an **iterative** algorithm based on optimal and sub-optimal score distributions from profile analysis that estimates the significance of all **repeats** that are detected in a single **sequence**. This procedure allows the identification of homologues at alignment scores lower than the highest optimal alignment score for non-homologous **sequences**. The method has been used to investigate the occurrence of eleven families of **repeats** in Saccharomyces cerevisiae, Caenorhabditis elegans and Homo sapiens accounting for 1055, 2205 and 2320 **repeats**, respectively. For these examples, the method is both more sensitive and more selective than conventional homology search procedures. The method allowed the detection in the SwissProt **database** of more than 2000 previously unrecognised **repeats** belonging to the 11 families. In addition, the method was used to merge several **repeat** families that previously were supposed to be distinct, indicating common phylogenetic origins for these families. Copyright 2000 Academic Press.

L5 ANSWER 18 OF 28 MEDLINE on STN DUPLICATE 10
AN 2001054028 MEDLINE
DN PubMed ID: 10966575
TI Rapid automatic detection and alignment of **repeats** in protein **sequences**.
AU Heger A; Holm L
CS European Bioinformatics Institute, Cambridge, United Kingdom.. heger@ebi.ac.uk
SO Proteins, (2000 Nov 1) 41 (2) 224-37.
Journal code: 8700181. ISSN: 0887-3585.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200012
ED Entered STN: 20010322
Last Updated on STN: 20010322
Entered Medline: 20001213
AB Many large proteins have evolved by internal duplication and many internal **sequence repeats** correspond to functional and structural units. We have developed an automatic algorithm, RADAR, for segmenting a query **sequence** into **repeats**. The segmentation procedure has three steps: (i) **repeat** length is determined by the spacing between suboptimal self-alignment traces; (ii) **repeat** borders are optimized to yield a maximal integer number of **repeats**, and (iii) distant **repeats** are validated by **iterative** profile alignment. The method identifies short composition biased as well

as gapped approximate **repeats** and complex **repeat** architectures involving many different types of **repeats** in the query **sequence**. No manual intervention and no prior assumptions on the number and length of **repeats** are required. Comparison to the Pfam-A **database** indicates good coverage, accurate alignments, and reasonable **repeat** borders. Screening the Swissprot **database** revealed 3,000 **repeats** not annotated in existing domain **databases**. A number of these **repeats** had been described in the literature but most were novel. This illustrates how in times when curated **databases** grapple with ever increasing backlogs, automatic (re)analysis of **sequences** provides an efficient way to capture this important information.

L5 ANSWER 19 OF 28 BIOSIS COPYRIGHT 2004 BIOLOGICAL ABSTRACTS INC. on STN
AN 2001:519197 BIOSIS
DN PREV200100519197
TI Strategies for building a physical map of human chromosome 15q14-15q21.3.
AU Madan, Anuradha [Reprint author]; Dors, Monica [Reprint author]; Qin, Shizhen [Reprint author]; Rowen, Lee [Reprint author]; Stewart, Sandra [Reprint author]; Madan, Anup [Reprint author]; Abbasi, Nissa [Reprint author]; Alexander, Shannon [Reprint author]; Bloom, Scott [Reprint author]; Birditt, Brian [Reprint author]; Baradarani, Lida [Reprint author]; Dickhoff, Rachel [Reprint author]; Fahey, Jessica [Reprint author]; Fleetwood, Peter [Reprint author]; Harrison, Grace [Reprint author]; Hicks, Jamie [Reprint author]; Johnson, Erica [Reprint author]; Kaur, Amardeep [Reprint author]; Shaffer, Tristan [Reprint author]; Friedman, Cynthia [Reprint author]; Hood, Leroy [Reprint author]
CS Institute for Systems Biology, Seattle, WA, USA
SO International Genome Sequencing and Analysis Conference, (2000) Vol. 12, pp. 76. print.
Meeting Info.: 12th International Genome Sequencing and Analysis Conference. Miami Beach, Florida, USA. September 12-15, 2000.
DT Conference; (Meeting)
Conference; Abstract; (Meeting Abstract)
Conference; (Meeting Poster)
LA English
ED Entered STN: 7 Nov 2001
Last Updated on STN: 23 Feb 2002
AB Our aim has been to produce long stretches of contiguous finished **sequence**. In less than a year, we generated a BAC-based physical map of human chromosome 15 (15q14-15q21.3) using a variety of wet lab and electronic strategies. These strategies included: Isolation of seed BAC clusters by hybridization with known STS, EST, or cDNA markers and validation of BACs in these clusters by FISH and restriction digest fingerprinting. These seed BAC clusters serve as a backbone of the map. Identification of minimally overlapping contig-extension and gap filling BACs suitable for draft sequencing by searching the BAC end (STC) **database** with the scaffolded **sequence** and/or by searching the St. Louis (FPC) fingerprint **database**. Identification of all BACs being sequenced in our target region by a) blasting all markers mapped to our target region against the HTGS **database** in GenBank, b) blasting BAC end **sequences** derived from BACs believed to overlap draft **sequences**, and c) blasting the **Repeat Masked** contigs of any draft **sequences** identified in the **database**. Construction of the complete physical map by analyzing the overlaps revealed by the **sequence** and fingerprint matches. Using these procedures our center has put together a physical map of chromosome 15 which is approximately 15-20 Mb in size.

L5 ANSWER 20 OF 28 BIOSIS COPYRIGHT 2004 BIOLOGICAL ABSTRACTS INC. on STN
AN 2001:494512 BIOSIS

DN PREV200100494512
TI EST assembly using Paracel's clustering package.
AU Boysen, Cecilie [Reprint author]; Borkowski, Joseph A. [Reprint author]; Candlin, James [Reprint author]; Herrmannsfeldt, Glen; Huang, Xiaoqiu [Reprint author]; Oh, KyungNa [Reprint author]; Paul, Cassi [Reprint author]; Qian, Jun [Reprint author]; Smith, Charles P. [Reprint author]; Xu, Licen [Reprint author]; Hunkapiller, Tim [Reprint author]
CS Paracel, Inc., Pasadena, CA, USA
SO International Genome Sequencing and Analysis Conference, (2000) Vol. 12, pp. 51-52. print.
Meeting Info.: 12th International Genome Sequencing and Analysis Conference. Miami Beach, Florida, USA. September 12-15, 2000.
DT Conference; (Meeting)
Conference; Abstract; (Meeting Abstract)
Conference; (Meeting Poster)
LA English
ED Entered STN: 24 Oct 2001
Last Updated on STN: 23 Feb 2002
AB We have developed a UNIX-based pipeline for clustering and assembly of DNA **sequences**. This provides a robust and flexible environment for EST clustering projects. The pipeline automatically converts **sequence** data from different **databases** into one common format. **Repeat sequences** as described by RepBase and low-complexity regions are **masked**. Contaminating **sequences** such as E. coli and mitochondria are removed. Paracel's clustering package can utilize a variety of algorithms for comparison of **sequences**. If full-length cDNAs or mRNAs, seeds, are available a pre-clustering step is performed, where each individual **sequence** is compared against the seeds and assigned to a seed-cluster if applicable. The remaining **sequences** go into pairwise comparison for clustering. Clustered ESTs are then assembled using CAP4. Alignments and consensus **sequences** can be viewed and edited using AssemblyView. All of the above can be performed with a single command which can be tailored to any specific set of ESTs. New data can be added and clustered and assembled onto existing projects. We will discuss timing and results obtained by using Paracel's clustering package on EST sets from various species.

L5 ANSWER 21 OF 28 MEDLINE on STN DUPLICATE 11
AN 1999455272 MEDLINE
DN PubMed ID: 10526352
TI An **iterative** structure-assisted approach to **sequence** alignment and comparative modeling.
AU Burke D F; Deane C M; Nagarajaram H A; Campillo N; Martin-Martinez M; Mendes J; Molina F; Perry J; Reddy B V; Soares C M; Steward R E; Williams M; Carrondo M A; Blundell T L; Mizuguchi K
CS Department of Biochemistry, University of Cambridge, United Kingdom.
SO Proteins, (1999) Suppl 3 55-60.
Journal code: 8700181. ISSN: 0887-3585.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199911
ED Entered STN: 20000111
Last Updated on STN: 20000111
Entered Medline: 19991109
AB Correct alignment of the **sequence** of a target protein with those of homologues of known three-dimensional structure is a key step in comparative modeling. Usually an **iterative** approach that takes account of the local and overall structural features is required. We describe such an approach that exploits **databases** of structural

alignments of homologous proteins (HOMSTRAD, http://(/)www-cryst.bioc.cam.ac.uk/ approximately homstrad) and protein superfamilies (CAMPASS, http://(/)www-cryst.bioc.cam.ac.uk/ approximately campass), in which structure-based alignments are analyzed and formatted with the program JOY (http://(/)www-cryst.bioc.cam.ac.uk/ approximately joy) to reveal conserved local structural features. The **databases** facilitate the recognition of a family or superfamily, they assist in the selection of useful parent structures, they are helpful in alignment of the target **sequences** with the parent set, and are useful for deriving relationships that can be used in validating models. In the **iterative** approach, a model is constructed on the basis of the proposed **sequence** alignment and this is then reexpressed in the JOY format and realigned with the parent set. This is **repeated** until the model and **sequence** alignment is optimized. We examine the case for comparison and use of multiple structures of family members, rather than a single parent structure. We use the targets attempted by our group in CASP3 to assess the value of such procedures.

L5 ANSWER 22 OF 28 MEDLINE on STN DUPLICATE 12
AN 1999265969 MEDLINE
DN PubMed ID: 10331942
TI Localization of retina/pineal-expressed **sequences**: identification of novel candidate genes for inherited retinal disorders.
AU Sohocki M M; Malone K A; Sullivan L S; Daiger S P
CS School of Public Health, The University of Texas Health Science Center, Houston, Texas 77225-0334, USA.
NC EY07024 (NEI)
EY07142 (NEI)
SO Genomics, (1999 May 15) 58 (1) 29-33.
Journal code: 8800135. ISSN: 0888-7543.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
OS GENBANK-G42173; GENBANK-G42174; GENBANK-G42175; GENBANK-G42176; GENBANK-G42177; GENBANK-G42178; GENBANK-G42179; GENBANK-G42180; GENBANK-G42181; GENBANK-G42182; GENBANK-G42183; GENBANK-G42184; GENBANK-G42185; GENBANK-G42186; GENBANK-G42187; GENBANK-G42188; GENBANK-G42189; GENBANK-G42190; GENBANK-G42191; GENBANK-G42192; GENBANK-G42193; GENBANK-G42194; GENBANK-G42195; GENBANK-G42196; GENBANK-G42197; GENBANK-G42198
EM 199907
ED Entered STN: 19990715
Last Updated on STN: 19990715
Entered Medline: 19990706
AB More than 100 genes causing inherited retinal diseases have been mapped to chromosomal locations, but less than half of these genes have been cloned. Mutations in many retina/pineal-specific genes are known to cause inherited retinal diseases. Examples include mutations in arrestin, rhodopsin kinase, and the cone-rod homeobox gene, CRX. To identify additional candidate genes for inherited retinal disorders, novel retina/pineal-expressed EST clusters were identified from the TIGR Human Gene Index **database** and mapped to specific chromosomal sites. After known human gene **sequences** were excluded, and **repeat sequences** were **masked**, 26 novel retina and pineal gland cDNA clusters were identified. The retinal expression of each novel EST cluster was confirmed by PCR assay of a retinal cDNA library, and each cluster was localized in the genome using the GeneBridge 4.0 radiation hybrid panel. In silico expression data from the TIGR **database** suggest that these EST clusters are retina/pineal-specific or predominantly expressed in these tissues. This combination of **database** analysis and laboratory investigation has localized

several EST clusters that are potential candidates for genes causing inherited retinopathy.
Copyright 1999 Academic Press.

L5 ANSWER 23 OF 28 MEDLINE on STN DUPLICATE 13
AN 1999170301 MEDLINE
DN PubMed ID: 10072085
TI **Iterative** stepwise discriminant analysis: a meta-algorithm for detecting quantitative **sequence** motifs.
AU Mallios R R
CS Medical Information Resources, University of California at San Francisco, Fresno 93703, USA.. ronna@ucsfresno.edu
SO Journal of computational biology : a journal of computational molecular cell biology, (1998 Winter) 5 (4) 703-11.
Journal code: 9433358. ISSN: 1066-5277.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199904
ED Entered STN: 19990517
Last Updated on STN: 19990517
Entered Medline: 19990430
AB An algorithm is presented for detecting a quantitative pattern in peptide fragments that bind class II major histocompatibility complex (MHC) molecules. It is referred to as a meta-algorithm because it requires successive applications of Stepwise Discriminate Analysis (SDA). On every iteration the best subsequence candidates are selected from **sequences** known to bind class II MHC molecules. When SDA compares probable binding subsequences with subsequences known not to bind class II MHC molecules, a quantitative model emerges that is capable of classifying subsequences as binding or non-binding. In an **iterative** manner, the resultant model is utilized as a criterion for selecting probable binding subsequence candidates. The procedure is **repeated** until models converge. In the illustrated examples, the final models correctly classify over 95% of the peptides in a **database** of peptides whose binding affinity for HLA-DR1 is known. The final model can then be used to predict the binding affinity of peptides that have not yet been laboratory tested.

L5 ANSWER 24 OF 28 MEDLINE on STN DUPLICATE 14
AN 1998349971 MEDLINE
DN PubMed ID: 9683596
TI Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes.
AU Krawczak M; Ball E V; Cooper D N
CS Institute of Medical Genetics, University of Wales College of Medicine,Cardiff CF4 4XN, United Kingdom.. krawczak@cardiff.ac.uk
SO American journal of human genetics, (1998 Aug) 63 (2) 474-88.
Journal code: 0370475. ISSN: 0002-9297.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199812
ED Entered STN: 19990115
Last Updated on STN: 19990115
Entered Medline: 19981224
AB The spectrum of single-base-pair substitutions logged in The Human Gene Mutation **Database** (HGMD), comprising 7,271 different lesions in the coding regions of 547 different human genes, was analyzed for nearest-neighbor effects on relative mutation rates. Owing to its

retrospective nature, HGMD allows mutation rates to be estimated only in relative terms. Therefore, a novel methodology was devised in order to obtain these estimates in **iterative** fashion, correcting, at the same time, for the confounding effects of differential codon usage and for the fact that different types of amino acid replacement come to clinical attention with different probabilities. Over and above the hypermutability of CpG dinucleotides, reflected in transition rates five times the base mutation rate, only a subtle and locally confined influence of the surrounding DNA **sequence** on relative single-base-pair substitution rates was observed, which extended no farther than 2 bp from the substitution site. A disparity between the two DNA strands was evidenced by the fact that, when substitution rates were estimated conditional on the 5' and 3' flanking nucleotides, a significant rate difference emerged for 10 of 96 possible pairs of complementary substitutional events. Mutational bias, favoring substitutions toward flanking bases, a phenomenon reminiscent of misalignment mutagenesis, was apparent and exhibited both directionality and reading-frame sensitivity. No specific preponderance of **repeat-sequence** motifs was observed in the vicinity of nucleotide substitutions, but a moderate correlation between the relative mutability and thermodynamic stability of DNA triplets emerged, suggesting either inefficient DNA replication in regions of high stability or the transient stabilization of misaligned intermediates.

L5 ANSWER 25 OF 28 MEDLINE on STN
AN 1998069599 MEDLINE
DN PubMed ID: 9406524
TI Identification of the components of simple protein mixtures by high-accuracy peptide mass mapping and **database** searching.
AU Jensen O N; Podtelejnikov A V; Mann M
CS Protein & Peptide Group, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany.. Jenseno@pr-group.ou.dk
SO Analytical chemistry, (1997 Dec 1) 69 (23) 4741-50.
Journal code: 0370536. ISSN: 0003-2700.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199802
ED Entered STN: 19980217
Last Updated on STN: 19980217
Entered Medline: 19980203
AB Peptide mass mapping by matrix-assisted laser desorption/ionization (MALDI) followed by **database** searching with the set of measured peptide masses is now a powerful method for the identification of pure proteins. Protein mixtures--such as frequently occur due to comigration in polyacrylamide gel bands--have hitherto required protein sequencing. Here we demonstrate that such protein bands can also be analyzed by peptide mass mapping alone. **Database** searching with the complete list of peptide masses determined by delayed-extraction MALDI mass spectrometry with a mass error of less than 30 ppm retrieves the most prominent protein in a mixture. In a second step, the protein identity is further confirmed by matching as many of the measured peptide masses as possible to the retrieved amino acid **sequence**. Peptide masses remaining after this "second pass search" are searched again to identify the next component in the protein mixture. This **iterative** process is **repeated** until all major ion signals are accounted for. Protein mixtures consisting of two or more individual components in a single gel band can be analyzed, further increasing the general applicability of MALDI peptide mapping for protein identification.

L5 ANSWER 26 OF 28 MEDLINE on STN

AN 97031596 MEDLINE
DN PubMed ID: 8877521
TI Compact encoding strategies for DNA **sequence** similarity search.
AU States D J; Agarwal P
CS Institute for Biomedical Computing, Washington University, St. Louis, MO 63110, USA.. states@ibc.wustl.edu
SO Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology, (1996) 4 211-7.
Journal code: 9509125.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199701
ED Entered STN: 19970219
Last Updated on STN: 19970219
Entered Medline: 19970130
AB Determining whether two DNA **sequences** are similar is an essential component of DNA **sequence** analysis. Dynamic programming is the algorithm of choice if computational time is not the most important consideration. Heuristic search tools, such as BLAST, are computationally more efficient, but they may miss some of the **sequence** similarities (Altschul et al., 1990). These tools often use common k-tuples (words) between the two **sequences** to determine anchor points for the alignment, and spend most of their computational time extending the alignment beyond these anchor points. We discuss and provide a DNA **sequence** similarity search implementation (called SENSEI) that improves upon the performance of BLASTN by almost an order of magnitude for comparable sensitivity. This improvement is a result of using compactly encoded scoring tables for k-tuples, encoding bases with a single bit, filtering the **sequence** to remove the simple **sequence repeats** using XNUN, and **masking** the known species-specific **repeats** in the query **sequence**. To reduce memory requirements, especially for large genomic DNA query **sequences**, we recommend generating the neighborhood words from the target **sequence** at run-time, instead of generating them by preprocessing the query **sequence**.

L5 ANSWER 27 OF 28 MEDLINE on STN DUPLICATE 15
AN 96348763 MEDLINE
DN PubMed ID: 8744771
TI XFINGER: a tool for searching and visualising protein fingerprints and patterns.
AU Perkins D N; Attwood T K
CS Department of Biochemistry, University of Leeds, UK.
SO Computer applications in the biosciences : CABIOS, (1996 Apr) 12 (2) 89-94.
Journal code: 8511758. ISSN: 0266-7061.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199611
ED Entered STN: 19961219
Last Updated on STN: 20000303
Entered Medline: 19961106
AB A tool for searching pattern and fingerprint **databases** is described. Fingerprints are groups of motifs excised from conserved regions of **sequence** alignments and used for **iterative database** scanning. The constituent motifs are thus encoded as small alignments in which **sequence** information is maximised with

each **database** pass; they therefore differ from regular-expression patterns, in which alignments are reduced to single consensus **sequences**. Different **database** formats have evolved to store these disparate types of information, namely the PROSITE dictionary of patterns and the PRINTS fingerprint **database**, but programs have not been available with the flexibility to search them both. We have developed a facility to do this: the system allows query **sequences** to be scanned against either PROSITE, the full PRINTS **database**, or against individual fingerprints. The results of fingerprint searches are displayed simultaneously in both text and graphical windows to render them more tangible to the user. Where structural coordinates are available, identified motifs may be visualised in a 3D context. The program runs on Silicon Graphics machines using GL graphics libraries and on machines with X servers supporting the PEX extension: its use is illustrated here by depicting the location of low-density lipoprotein-binding (LDL) motifs and leucine-rich **repeats** in a mosaic G-protein-coupled receptor (GPCR).

L5 ANSWER 28 OF 28 MEDLINE on STN DUPLICATE 16
AN 96267105 MEDLINE
DN PubMed ID: 8654539
TI Plasmodium vivax: favored gene frequencies of the merozoite surface protein-1 and the multiplicity of infection in a malaria endemic region.
AU Kolakovich K A; Ssengoba A; Wojcik K; Tsuboi T; al-Yaman F; Alpers M; Adams J H
CS Department of Biological Sciences, University of Notre Dame, Indiana 46556, USA.
NC R29 AI33656 (NIAID)
SO Experimental parasitology, (1996 Jun) 83 (1) 11-9.
Journal code: 0370713. ISSN: 0014-4894.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
OS GENBANK-U31924; GENBANK-U31925; GENBANK-U31926; GENBANK-U31927; GENBANK-U31928
EM 199608
ED Entered STN: 19960808
Last Updated on STN: 19990129
Entered Medline: 19960801
AB In this study, we present an analysis of the Plasmodium vivax MSP-1 polymorphic region 5 and identify a new recombinant gene element. In clinical isolates from Papua New Guinea (PNG), the P. vivax MSP-1 gene type was characterized by restriction fragment length polymorphisms and by Southern blot oligonucleotide hybridizations using probes to type-specific **sequences**. There were three pairs of dimorphic gene elements in the MSP-1 polymorphic region 5; four of the eight potential different combinations of **sequence** elements for this region have been identified. The center gene segment was the most polymorphic, especially for the glutamine (Q) **repeat** element with virtually every gene containing a different length of Q **repeats**, a finding consistent with **database sequence** information. The frequencies of all of the polymorphic MSP-1 gene elements were approximately equal except for the first segment, which was biased 10:1 for the Type II (Sal-1 type) versus Type I (Belem type) gene segment. In fact, only one combination (I/Q/S) of the genetic elements containing the type I gene segment for polymorphic region 5 was identified, a finding consistent with **sequences** reported to gene data banks. Considering only the multiplicity of MSP-1 gene types, 38% of the patients were identified as having multiple infections; when correlated with the circumsporozoite protein and the Duffy antigen binding protein gene types, the multiple infection rate increased to 65% of 23 isolates characterized. Increased

age was the only clinical parameter that positively correlated with multiclonal infections and there was no other apparent bias or linkage of gene types among the three loci. These data identify multiple clonal populations of P. vivax in the PNG population and potentially a high rate of concurrent infections in clinical cases. The extreme polymorphism of the MSP-1 polymorphic region 5 suggests that frequent recombination occurs within this gene. The bias in frequency for one recombinant gene motif indicates that intrinsic host or parasite factors may engender increased frequency of one genetic element over another. Failure to identify this type of discrete clonal marker as well as reliance on a single marker can **mask** the true multiclonal nature of an infection and lead to underestimation of the multiplicity of infection.

WEST Search History

Hide Items | Restore | Clear | Cancel

DATE: Tuesday, July 20, 2004

Hide?	Set Name	Query	Hit Count
		DB=PGPB,USPT,EPAB,JPAB,DWPI; PLUR=YES; OP=ADJ	
☐	L3	2000	18
☐	L2	L1 same (iterative or mask)	89
☐	L1	sequence same database same repeat$	2429

END OF SEARCH HISTORY

Hit List

Clear | Generate Collection | Print | Fwd Refs | Bkwd Refs | Generate OACS

Search Results - Record(s) 1 through 18 of 18 returned.

☐ 1. Document ID: US 20030077692 A1

L3: Entry 1 of 18 File: PGPB Apr 24, 2003

PGPUB-DOCUMENT-NUMBER: 20030077692
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030077692 A1

TITLE: REFOLDING METHOD

PUBLICATION-DATE: April 24, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
FERSHT, ALAN ROY	CAMBRIDGE		GB	
ALTAMIRANO, MYRIAM MARLENNE	CAMBRIDGE		GB	

US-CL-CURRENT: 435/68.1

ABSTRACT:

The invention relates to a method for promoting the folding of a polypeptide, comprising the step of contacting the polypeptide with a molecular chaperone and a foldase.

L3: Entry 1 of 18 File: PGPB Apr 24, 2003

DOCUMENT-IDENTIFIER: US 20030077692 A1
TITLE: REFOLDING METHOD

Application Filing Year:
1999

Detail Description Paragraph:
[0138] FILTER Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (1993) Computers and Chemistry 17:149-163, or segments consisting of short-periodicity internal repeats, as determined by the XNU program of Claverie & States (1993) Computers and Chemistry 17:191-201, or, for BLASTN, by the DUST program of Tatusov and Lipman (see http://www.ncbi.nlm.nih.gov). Filtering can eliminate statistically significant but biologically uninteresting reports from the blast output (e.g., hits against common acidic-, basic- or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.

☐ 2. Document ID: US 20020168744 A1

L3: Entry 2 of 18 File: PGPB Nov 14, 2002

PGPUB-DOCUMENT-NUMBER: 20020168744
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20020168744 A1

TITLE: AMINO ACID SEQUENCE

PUBLICATION-DATE: November 14, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
BRUNSTEDT, JANNE	ROSKILDE		DK	
CHRISTENSEN, TOVE MARTEL IDA ELSE	ALLEROD		DK	

US-CL-CURRENT: 435/197; 426/50, 435/101, 435/275, 435/440, 536/23.2

ABSTRACT:

An amino acid sequence is described that affects PME activity. The amino acid has the formula (I):

A1-A2-A3-A4-A5-A6-A7-A8-A9-A10-A11-A12-A13-A14-A15-A16-A17-A18-A19-A20-A21- -A22 (I)

L3: Entry 2 of 18 File: PGPB Nov 14, 2002

DOCUMENT-IDENTIFIER: US 20020168744 A1
TITLE: AMINO ACID SEQUENCE

Application Filing Year:
1999

Summary of Invention Paragraph:
[0255] FILTER Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (1993) Computers and Chemistry 17:149-163, or segments consisting of short-periodicity internal repeats, as determined by the XNU program of Claverie & States (1993) Computers and Chemistry 17:191-201, or, for BLASTN, by the DUST program of Tatusov and Lipman (see http://www.ncbi.nlm.nih.gov). Filtering can eliminate statistically significant but biologically uninteresting reports from the blast output (e.g., hits against common acidic-, basic- or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.

☐ 3. Document ID: US 20020115176 A1

L3: Entry 3 of 18 File: PGPB Aug 22, 2002

PGPUB-DOCUMENT-NUMBER: 20020115176
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20020115176 A1

TITLE: PHOSPHODIESTERASE ENZYMES

PUBLICATION-DATE: August 22, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
LANFEAR, JEREMY	SANDWICH		GB	
ROBAS, NICOLA M.	SANDWICH		GB	

US-CL-CURRENT: 435/196; 424/551, 435/19, 435/320.1, 435/325, 514/44, 530/387.1, 536/23.2, 536/23.5

ABSTRACT:

This invention provides novel PDE11 genes and polypeptides, and variants, homologues, fragments, and derivatives thereof. This invention also provides vectors and host cells comprising the disclosed nucleotide sequences. This invention further provides antibodies that bind to the PDE11 polypeptides. This invention further yet provides methods for identifying agents that affect the expression or activity of the PDE11 genes and polypeptides. This invention also provides pharmaceutical compositions comprising the PDE11 genes or polypeptides, or inhibitors thereof. This invention additionally provides methods for treating diseases and conditions related to PDE11 activity, or the inhibition thereof.

L3: Entry 3 of 18 File: PGPB Aug 22, 2002

DOCUMENT-IDENTIFIER: US 20020115176 A1
TITLE: PHOSPHODIESTERASE ENZYMES

Application Filing Year:
1999

Detail Description Paragraph:
[0353] FILTER. Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (1993) Computers and Chemistry 17:149-163, or segments consisting of short-periodicity internal repeats, as determined by the XNU program of Clayerie & States (1993) Computers and Chemistry 17:191-201, or, for BLASTN, by the DUST program of Tatusov and Lipman (see http://www.ncbi.nlm.nih.gov). Filtering can eliminate statistically significant but biologically uninteresting reports from the blast output (e.g., hits against common acidic-, basic- or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.

☐ 4. Document ID: US 20020102539 A1

L3: Entry 4 of 18 File: PGPB Aug 1, 2002

PGPUB-DOCUMENT-NUMBER: 20020102539
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20020102539 A1

TITLE: NUCLEOTIDE SEQUENCES AND PROTEIN SEQUENCES

PUBLICATION-DATE: August 1, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
ARKOWITZ, ROBERT ALAN	CAMBRIDGE		GB	
NERN, PETER MICHAEL ALJOSCHA	CAMBRIDGE		GB	

US-CL-CURRENT: 435/6; 435/91.2, 536/23.1

ABSTRACT:

A nucleotide sequence is described. The nucleotide sequence or the expression product of the nucleotide sequence has the capability of not substantially affecting the interaction of G.beta. with Cdc24p or a hornologue thereof that is usually capable of being associated with the Cdc24p or the homologue thereof.

L3: Entry 4 of 18 File: PGPB Aug 1, 2002

DOCUMENT-IDENTIFIER: US 20020102539 A1
TITLE: NUCLEOTIDE SEQUENCES AND PROTEIN SEQUENCES

Application Filing Year:
1998

Summary of Invention Paragraph:
[0085] FILTER Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (1993) Computers and Chemistry 17:149-163, or segments consisting of short-periodicity internal repeats, as determined by the XNU prograrn of Claverie & States (1993) Computers and Chemistry 17:191-201, or, for BLASTN, by the DUST program of Tatusov and Lipman (see http://www.ncbi.nlm.nih.gov). Filtering can ellimnate statistically significant but biologically uninteresting reports from the blast output (e.g., hits against common acidic-, basic-or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.

Full | Title | Citation | Front | Review | Classification | Date | Reference | Sequences | Attachments | Claims | KWIC | Draw D

☐ 5. Document ID: US 20020099169 A1

L3: Entry 5 of 18 File: PGPB Jul 25, 2002

PGPUB-DOCUMENT-NUMBER: 20020099169
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20020099169 A1

TITLE: TPL-2/COT KINASE AND METHODS OF USE

PUBLICATION-DATE: July 25, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
ALLEN, HAMISH JOHN	BOYLSTON	MA	US	
DIXON, RICHARD WOODWARD	NORTH GRAFTON	MA	US	
KAMENS, JOANNE SARA	NEWTON CENTRE	MA	US	
WICKRAMASINGHE, DINELI	NEWTON	MA	US	
XU, YAJUN	WESTBOROUGH	MA	US	
BELICH, MONICA POLIDORO	LONDON		GB	
JOHNSTON, LELAND HERRIES	LONDON		GB	
LEY, STEVEN CHARLES	HIGH BARNET		GB	
SALMERON, ANDRES	LONDON		GB	

US-CL-CURRENT: 530/324; 424/130.1, 435/7.1, 530/350, 530/387.1

ABSTRACT:

It is shown that TPL-2 is responsible for phosphorylation of p105 and its resultant proteolysis, which leads to p50 Rel translocation to the nucleus. Accordingly, the invention provides TPL-2 as a specific regulator of the activation of NF.kappa.B, and thus as a modulator of inflammatory responses in which p105 is involved, and as a target for the development of compounds capable of influencing NF.kappa.B activation.

L3: Entry 5 of 18 File: PGPB Jul 25, 2002

DOCUMENT-IDENTIFIER: US 20020099169 A1
TITLE: TPL-2/COT KINASE AND METHODS OF USE

Application Filing Year:
1999

Detail Description Paragraph:
[0104] FILTER Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (1993) Computers and Chemistry 17:149-163, or segments consisting of short-periodicity internal repeats, as determined by the XNU program of Claverie & States (1993) Computers and Chemistry 17:191-201, or, for BLASTN, by the DUST program of Tatusov and Lipman (see http://www.nchi.nlm.nih.gov). Filtering can eliminate statistically significant but biologically uninteresting reports from the blast output (e.g., hits against common acidic-, basic- or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.

☐ 6. Document ID: US 6743609 B1

L3: Entry 6 of 18 File: USPT Jun 1, 2004

US-PAT-NO: 6743609
DOCUMENT-IDENTIFIER: US 6743609 B1

TITLE: Linoleate isomerase

DATE-ISSUED: June 1, 2004

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Rosson; Reinhardt A.	Manitowoc	WI		
Grund; Alan D.	Manitowoc	WI		
Deng; Ming-De	Manitowoc	WI		
Sanchez-Riera; Fernando	Manitowoc	WI		

US-CL-CURRENT: 435/134; 435/176, 435/177, 435/233

ABSTRACT:

The present invention provides an isolated linoleate isomerase and its nucleic acid and amino acid sequence. The present invention also provides a method for producing CLA from an oil using an immobilized bacterial cell or an isolated linoleate isomerase.

47 Claims, 47 Drawing figures
Exemplary Claim Number: 1
Number of Drawing Sheets: 42

L3: Entry 6 of 18 File: USPT Jun 1, 2004

DOCUMENT-IDENTIFIER: US 6743609 B1
TITLE: Linoleate isomerase

Application Filing Year (1):
1998

Detailed Description Paragraph Table (1):
TABLE 1 BLAST Search Parameters HISTOGRAM Display a histogram of scores for each search; default is yes. (See parameter H in the BLAST Manual). DESCRIPTIONS Restricts the number of short descriptions of matching sequences reported to the number specified; default limit is 100 descriptions. (See parameter V in the manual page). See also EXPECT and CUTOFF. ALIGNMENTS Restricts database sequences to the number specified for which high-scoring segment pairs (HSPs) are reported; the default limit is 50. If more database sequences than this happen to satisfy the statistical significance threshold for reporting (see EXPECT and CUTOFF below), only the matches ascribed the greatest statistical significance are reported. (See parameter B in the BLAST Manual). EXPECT The statistical significance threshold for

reporting matches against database sequences; the default value is 10, such that 10 matches are expected to be found merely by chance, according to the stochastic model of Karlin and Altschul (1990). If the statistical significance ascribed to a match is greater than the EXPECT threshold, the match will not be reported. Lower EXPECT thresholds are more stringent, leading to fewer chance matches being reported. Fractional values are acceptable. (See parameter E in the BLAST Manual). CUTOFF Cutoff score for reporting high-scoring segment pairs. The default value is calculated from the EXPECT value (see above). HSPs are reported for a database sequence only if the statistical significance ascribed to them is at least as high as would be ascribed to a lone HSP having a score equal to the CUTOFF value. Higher CUTOFF values are more stringent, leading to fewer chance matches being reported. (See parameter S in the BLAST Manual). Typically, significance thresholds can be more intuitively managed using EXPECT. MATRIX Specify an alternate scoring matrix for BLASTP, BLASTX, TBLASTN and TBLASTX. The default matrix is BLOSUM62 (Henikoff & Henikoff, 1992). The valid alternative choices include: PAM40, PAM120, PAM250 and IDENTITY. No alternate scoring matrices are available for BLASTN; specifying the MATRIX directive in BLASTN requests returns an error response. STRAND Restrict a TBLASTN search to just the top or bottom strand of the database sequences; or restrict BLASTN, BLASTX or TBLASTX search to just reading frames on the top or bottom strand of the query sequence. FILTER Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (Computers and Chemistry, 1993), or segments consisting of short-periodicity internal repeats, as determined by the SNU program of Claverie & States (Computers and Chemistry, 1993), or, for BLASTN, by the DUST program of Tatusov and Lipman (in preparation). Filtering can eliminate statistically significant but biologically uninteresting reports from the blast output (e.g., hits against common acidic-, basic- or proline- rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences. Low complexity sequence found by a filter program is substituted using the letter "N" in nucleotide sequence (e.g., "NNNNNNNNNNNNN") and the letter "X" in protein sequences (e.g., "XXXXXXXXX"). Users may turn off filtering by using the "Filter" option on the "Advanced options for the BLAST server" page. Filtering is only applied to the query sequence (or its translation products), not to database sequences. Default filtering is DUST for BLASTN, SEG for other programs. It is not unusual for nothing at all to be masked by SEG, SNU, or both, when applied to sequences in SWISS-PROT, so filtering should not be expected to always yield an effect. Furthermore, in some cases, sequences are masked in their entirety, indicating that the statistical significance of any matches reported against the unfiltered query sequence should be suspect. NCBI-gi Causes NCBI gi identifiers to be shown in the output, in addition to the accession and/or locus name.

☐ 7. Document ID: US 6498020 B1

L3: Entry 7 of 18 File: USPT Dec 24, 2002

US-PAT-NO: 6498020
DOCUMENT-IDENTIFIER: US 6498020 B1

TITLE: Fusion proteins comprising coiled-coil structures derived of bovine IF1 ATPase inhibitor protein

DATE-ISSUED: December 24, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Walker; John	Cambridge			GB
Miroux; Bruno	Cambridge			GB

US-CL-CURRENT: 435/69.1; 424/184.1, 424/185.1, 435/69.7, 530/350, 530/412, 530/413

ABSTRACT:

The present invention relates to a fusion protein comprising a first amino acid sequence comprising the sequence of the C-terminal 40 amino acids of bovine IF.sub.1 ATPase inhibitor protein, and a second amino acid sequence not naturally associated with the first region. The invention further relates to methods for preparing an immunoglobulin comprising immunizing an animal with the fusion protein and recovering immunoglobulin specific for a region of the fusion protein.

9 Claims, 0 Drawing figures
Exemplary Claim Number: 1

L3: Entry 7 of 18 File: USPT Dec 24, 2002

DOCUMENT-IDENTIFIER: US 6498020 B1
TITLE: Fusion proteins comprising coiled-coil structures derived of bovine IF1 ATPase inhibitor protein

Application Filing Year (1):
1999

Brief Summary Text (27):
BLAST uses the following search parameters: HISTOGRAM Display a histogram of scores for each search; default is yes. (See parameter H in the BLAST Manual). DESCRIPTIONS Restricts the number of short descriptions of matching sequences reported to the number specified; default limit is 100 descriptions. (See parameter V in the manual page). See also EXPECT and CUTOFF. ALIGNMENTS Restricts database sequences to the number specified for which high-scoring segment pairs (HSPs) are reported; the default limit is 50. If more database sequences than this happen to satisfy the statistical significance threshold for reporting (see EXPECT and CUTOFF below), only the matches ascribed the greatest statistical significance are reported. (See parameter B in the BLAST Manual). EXPECT The statistical significance threshold for reporting matches against database sequences; the default value is 10, such that 10 matches are expected to be found merely by chance, according to the stochastic model of Karlin and Altschul (1990). If the statistical significance ascribed to a match is greater than the EXPECT threshold, the match will not be reported. Lower EXPECT thresholds are more stringent, leading to fewer chance matches being reported. Fractional values are acceptable. (See parameter E in the BLAST Manual). CUTOFF Cutoff score for reporting high-scoring segment pairs. The default value is calculated from the EXPECT value (see above). HSPs are reported for a database sequence only if the statistical significance ascribed to them is at least as high as would be ascribed to a lone HSP having a score equal to the CUTOFF value. Higher CUTOFF values are more stringent, leading to fewer chance matches being reported. (See parameter S in the BLAST Manual). Typically, significance thresholds can be more intuitively managed using EXPECT. MATRIX Specify an alternate scoring matrix for BLASTP, BLASTX. TBLASTN and TBLASTX. The default matrix is BLOSUM62 (Henikoff & Henikoff, 1992). The valid alternative choices include: PAM40, PAM120, PAM250 and IDENTITY. No alternate scoring matrices are available for BLASTN; specifying the MATRIX directive in BLASTN requests returns an error response. STRAND Restrict a TBLASTN search to just the top or

bottom strand of the database sequences; or restrict a BLASTN, BLASTX or TBLASTX search to just reading frames on the top or bottom strand of the query sequence. FILTER Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (1993) Computers and Chemistry 17:149-163, or segments consisting of short-periodicity internal repeats, as determined by the XNU program of Claverie & States (1993) Computers and Chemistry 17:191-201, or, for BLASTN, by the DUST program of Tatusov and Lipman (see http://www.ncbi.nlm.nih.gov). Filtering can eliminate statistically significant but biologically uninteresting reports from the blast output (e.g., hits against common acidic-, basic- or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.

☐ 8. Document ID: US 6496022 B1

L3: Entry 8 of 18 File: USPT Dec 17, 2002

US-PAT-NO: 6496022
DOCUMENT-IDENTIFIER: US 6496022 B1
**** See image for Certificate of Correction ****

TITLE: Method and apparatus for reverse engineering integrated circuits by monitoring optical emission

DATE-ISSUED: December 17, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Kash; Jeffrey A.	Pleasantville	NY		
Tsang; James C.	White Plains	NY		
Knebel; Daniel R.	Carmel	NY		

US-CL-CURRENT: 324/752

ABSTRACT:

A method and apparatus for reverse engineering an integrated circuit chip (IC chip) (120) utilizes an electrical circuit tester (114) for injecting a triggering signal into the IC chip (120) to exercise a circuit under test. In synchronization thereto, a PICA detector (116) monitors optical emissions from the circuit under test. A spatial data extractor, electrically coupled to the PICA detector, collects space information (124) from patterns of light emissions emitted by the circuit under test, and a timing data extractor, electrically coupled to the electrical circuit tester and to the PICA detector (116), collects time information (126) from the patterns of light emissions emitted by the circuit under test. A database memory (105) includes known data about the circuit under test and also includes at least one reference pattern for comparing a captured light emission pattern thereto to identify at least one circuit element in the circuit under test. A PICA data analyzer (108), electrically coupled to the database memory (105) and to the PICA detector (116), determines at least one of whether the circuit under test comprises a circuit element with a light emission pattern that matches one of the at least one reference pattern in the database memory (105), and the value contained in a

memory in the IC chip (120).

8 Claims, 3 Drawing figures
Exemplary Claim Number: 1
Number of Drawing Sheets: 3

L3: Entry 8 of 18 File: USPT Dec 17, 2002

DOCUMENT-IDENTIFIER: US 6496022 B1
**** See image for Certificate of Correction ****
TITLE: Method and apparatus for reverse engineering integrated circuits by monitoring optical emission

Application Filing Year (1):
1999

Detailed Description Text (20):
We will categorize reverse engineering into four classes as follows: (a) determining the physical locations of the subcircuits or circuit elements comprising the chip including the contents of various types of static memory circuits, (b) determining the logical functions and other functional characteristics of the subcircuits or circuit elements comprising the chip, (c) determining the device-level schematic of the transistors comprising each subcircuit or circuit element, and (d) determining the performance of the subcircuits or circuit elements). Specific examples are given below. (a) Determining the physical locations of the subcircuits or circuit elements comprising the chip In an alternative exemplary method for reverse engineering a circuit in an IC chip, the locations of scan chain latch elements can be determined by operating the scan chains in flush mode, to see which circuit elements of the IC are active. The active elements will be readily identified by the presence of emission. In an iterative procedure, as in the flow chart of FIG. 2, after this first identification, the scan chains can then be operated in a clocked mode. The additional circuit elements with produce emission are then those related to the scan clock circuitry. In designs that are found to not have a mode for flushing through the scan chains, the same net result can be obtained by first loading the scan chains with all zeros, applying zero to the scan inputs, and repeatedly scan clocking the latches while storing photon emission data, followed by loading the scan chains with alternating zeros and ones and repeatedly scanning alternating zeros and ones through the scan chains while storing photon emission data. Comparison of the two stored data files would reveal the clock circuits as emission patterns which did not change, and the scan latches as emission patterns which did change. In another alternative exemplary method for reverse engineering a circuit in an IC Chip, the reverse engineering system can determine a clock signal distribution network across the IC Chip. This method is useful, for example, for determining major logic blocks within a chip that are usually all linked to a common clock signal. Most IC Chips have publicly available test vectors for powering and exercising the clock circuit for the IC Chip. This is a commonly available test vector to circuit designers. Once the clock power circuit is exercised by the circuit tester, the PICA system can monitor light emissions from across the IC Chip to identify the location of timing circuit elements across the IC Chip. In a similar method to the above methods to identify the scan chain circuitry and clock circuit network, the connectivity of circuit elements to other circuit elements can be determined by exercising a target circuit element and seeing which other circuit elements are also active. Furthermore, by time-ordering the emission pulses from the various connected circuit elements, it is possible to determine the progression of the connections from one circuit element to the next. In the above example of the scan chain operated in flush mode, the time ordering of the emission pulses from each scan latch determines unambiguously the ordering of

the scan latches. In the case of the clock network, the time order determines the topology of the clock circuitry across the entire chip. (b) Determining the logical functions and other functional characteristics of the subcircuits or circuit elements comprising the chip. In an alternative embodiment of the present invention, as illustrated in FIG. 3, the reverse engineering system can be used to reverse engineer the contents of a static memory device such as a read only memory. A memory read out circuit, for example, can be repeatedly exercised to read out the value of a memory cell. A test vector can be repeatedly executed by a circuit tester 314 to continuously and repeatedly read out the value of a memory cell in an IC. The read out control circuit, in response to repeatedly reading out the value of a memory cell, repeatedly emits a pattern of light emissions that can be collected by the PICA system 316 to capture a profile of the read output of the memory cell. For example, the PICA system can determine the read output of a ROM cell. This creates a profile of the contents, or value, of the ROM cell by monitoring the light emissions therefrom during repeated read cycling of the output circuits of the ROM cell. The light emissions are collected with the PICA system 316 that is time synchronized to the circuit tester. The PICA system 316 in this way measures and profiles the wave forms from the ROM read out buffer. If the design of the memory cell read out buffer is known and preferably can be exercised, then one can simulate what optical wave form would be expected for a ROM cell value equal to zero and similarly what optical wave form would be expected for a ROM cell value equal to one. Typically, a one to zero transition at the output of a readout buffer will produce a much larger pulse of optical emissions than a zero to one transition. By monitoring these transitions relative to a known time base the reverse engineering system can determine the value stored in the ROM. The reverse engineering system 102 would compute both simulations for zero-to-one and for one-to-zero transitions and would have them stored in a database as known profiles or templates. Then, the reverse engineering system would compare them to the "unknown" measured profile to determine which simulation matched a best fit to the pattern in the measured profile. The result 328 then would indicate whether a ROM cell was at the value of zero or at a value of one. A discussion of a method for using a PICA system to deduce the value stored in a five bit counter was published in July, 1997 in a journal entitled "Electron Device Letters", in an article entitled "Dynamic Internal Testing of CMOS Circuits Using Hot Luminescence" by some of the inventors of the present invention. This published method by the present inventors is not an example of reverse engineering since the circuit design was already known. It is, however, an exemplary illustration of how PICA can be used to deduce a value temporarily stored in a counter circuit. FIG. 3 is a method outlining how to deduce a value permanently stored in a ROM circuit. For example if a reverse engineering system did not know how to exercise a ROM device to read out a value from its buffer, the reverse engineering system might first apply a reverse engineering method as discussed above to determine the detailed information of devices in the readout buffer portion of the ROM device. Then, once the readout buffer is characterized and the circuit elements are determined and essentially "known" by the reverse engineering system, this known information can be used to exercise the readout control circuit of the readout buffer of the ROM to determine the value contained in the ROM. The reverse engineering process, therefore, can be an iterative process to progressively determine additional information about a circuit under test. Other functional characteristics of a device may be reverse engineered in a similar manner. Examples include the ability to determine the sequence of operations taken to achieve a particular result. In a simple case it may be known that several operations and a particular number of clock cycles (or states in a state machine) are needed to achieve a particular result. The reverse engineering procedures described here may be used to determine the apportionment of the total number of clock cycles between the various operations needed to achieve the result. More specific examples follow. Given a device that has already been reverse engineered and it is understood that a particular circuit performs an addition of two numbers, analysis of the light emission from the elements of the adder circuit over a period of time that includes several addition operations would reveal whether one add operation must be completed before another is started. By applying

successive add operations to the device, measuring the location and time of light emitted from the adder elements, and comparing the start of one add operation to the end of the previous add operations, one could detect how many clock cycles were required to complete a single operation and how many add operations were started before completion of the first add operation (execution pipelining), or how many add operations were started between successive clock operations (wave pipelining). In a similar fashion, if it were known that both an adder circuit and a divider circuit were on the device, then analysis of the time-domain results of the light emission could be used to determine if only one of these circuits could operate at one time or if both could be made to operate in parallel. This information is useful to quickly determine implementation details of a complex circuit that may not be easily determined from the circuit topologies. Examples of such implementation details include but are not limited to multiple operation dispatch and vector and super scalar implementations. It is also useful to understand which operations on a chip are synchronized or not synchronized with other operations. One such example is cycle stealing, where a clock signal is delayed to one or more storage elements so that more computation may be achieved between clock cycle boundaries. Another example is an operation that completes in a multiple number of clock cycles without capturing the intermediate states of the operation in storage elements. (c) Determining the device-level schematic of the transistors comprising each subcircuit For example, as a simple case, suppose an IC chip comprises a delay circuit, which includes an unknown number of inverters in an inverter chain. In this example, it may be desired to determine whether there is an even number or an odd number of inverters, and how many stages are in the delay circuit.
Additionally, suppose one knew in advance how to inject a signal into this IC chip so that it would then propagate through this chain of inverters. Then, one could exercise the chain of inverters by propagating a signal through the chain, and by counting the subcircuits seen in the emission image directly determine the number of inverters in the chain. Similarly, if a divide-by-n circuit was found on an IC, with n unknown, one can determine the value of n by time-resolving the emission from the circuit. The value of n is then the frequency of the emission pulses at the input to the circuit divided by the frequency at the output. Note that for this reverse engineering situation, the time resolution of the emission is essential.
(d) Determining the performance of the subcircuits or circuit elements It is often useful to determine the performance of subcircuits as part of reverse engineering, so as to determine the ultimate capabilities of the circuit, such as speed, tolerance environmental conditions such as high temperature, and radio frequency interference immunity. Returning again to the scan chain operated in flush mode, or a chain of inverters, by time-resolving the emission from the sequential scan latches or inverters, one can directly measure the latch-to-latch or inverter-to-inverter delay. (The measurement of inverter-to-inverter delays is disclosed in an article entitled "Dynamic Internal Testing of CMOS Circuits Using Hot Luminescence" by some of the inventors of the present invention.) An advantage of using time resolved emission over the conventional method, which is to simply measure the delay between the first and last element of the chain, is that the present method allows measurement of individual delays, instead of just the average delay, so that the variations around the average, as well as the average, can be seen. Since the ultimate speed of operation of the chain is determined by the slowest element, the ability to see the individual delays can be a significant improvement in reverse engineering the circuit as compared to measuring only the average. Measurement such as those described in the previous paragraph can be made as a function of temperature or in the presence of strong radio frequency interference so as to determine the sensitivity of the individual circuit elements to these or other environmental influences.

☐ 9. Document ID: US 6340577 B1

L3: Entry 9 of 18 File: USPT Jan 22, 2002

US-PAT-NO: 6340577
DOCUMENT-IDENTIFIER: US 6340577 B1

TITLE: Protein fragments for use in protein targeting

DATE-ISSUED: January 22, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Hope; Ralph Graham	Glasgow			GB
McLauchlan; John	Glasgow			GB

US-CL-CURRENT: 435/69.7; 435/5, 530/326, 530/330, 530/387.3

ABSTRACT:

A protein is described. The protein comprises a lipid globule targeting sequence linked to a protein of interest (POI) wherein the targeting sequence comprises a hepatitis C virus (HCV) core protein or fragment or homologue thereof.

12 Claims, 63 Drawing figures
Exemplary Claim Number: 1
Number of Drawing Sheets: 22

L3: Entry 9 of 18 File: USPT Jan 22, 2002

DOCUMENT-IDENTIFIER: US 6340577 B1
TITLE: Protein fragments for use in protein targeting

Application Filing Year (1):
1998

Brief Summary Text (51):
FILTER Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (1993), or segments consisting of short-periodicity internal repeats, as determined by the XNU program of Claverie & States (1993), or, for BLASTN, by the DUST program of Tatusov and Lipman (see http://www.ncbi.nlm.nih.gov). Filtering can eliminate statistically significant but biologically uninteresting reports from the blast output (e.g. hits against common acidic-, basic- or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.

☐ 10. Document ID: US 6299649 B1

L3: Entry 10 of 18 File: USPT Oct 9, 2001

US-PAT-NO: 6299649
DOCUMENT-IDENTIFIER: US 6299649 B1

TITLE: Unbalanced prosthetic device for providing side-dependent twisting-rotational axial-loading coupling

DATE-ISSUED: October 9, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Chang; Fu-Kuo	Stanford	CA	94305	
Yildiz; Hasan	No. 19 G. O. P. Istanbul			TR
Goodman; Stuart B.	Los Altos	CA	94022	

US-CL-CURRENT: 623/23.34; 623/23.44, 623/23.51

ABSTRACT:

The present invention discloses a side-dependent prosthetic device. The side-dependent prosthetic device is a composite prosthetic device with a longitudinal direction. The composite prosthetic device includes a plurality of plies wherein each ply being composed of a plurality of reinforced fibers aligned in a ply orientational angle .theta.i relative to the longitudinal direction of the prosthetic device, where i=1,2,3, . . . ,N and N being the number of the plies. The plurality of plies are laminated together for forming the prosthetic device wherein the ply orientational angles being arranged such that .theta.1+.theta.2+.theta.3+ . . . +.theta..sub.N -0 thus forming an unbalanced composite prosthetic device. In another preferred embodiment, the ply orientational angles .theta.i forming a sequence which is represented by ((.theta..sup.1 /.theta..sup.2).sub.s).sub.n where a first ply with orientational angle .theta..sup.1 being followed by a second ply with orientational angle .theta..sup.2 wherein such a sequence repeated n-times and arranged to be symmetrical to a mid-plane, wherein .theta..sup.1 being an angle close to -10.sub.i and .theta..sup.2 being an angle close to -20.sub.i.

5 Claims, 21 Drawing figures
Exemplary Claim Number: 1
Number of Drawing Sheets: 11

L3: Entry 10 of 18 File: USPT Oct 9, 2001

DOCUMENT-IDENTIFIER: US 6299649 B1
TITLE: Unbalanced prosthetic device for providing side-dependent twisting-rotational axial-loading coupling

Application Filing Year (1):
1999

Detailed Description Text (4):
A flow chart is shown in FIG. 7 to illustrate a design process applying the stack sequence of ply-orientations as a design parameters. A prosthetic device with specific performance characteristics of stiffness, stress shielding, and micromotion can be obtained by employing the anisotropic nature of the ply orientations. The design process begins (step 200) with receiving patient's database (step 210) including the data for geometrical configuration of the bone, relative position of the bone in the body, the body weight, and data relating to

stiffness, loading, mechanical movements, force transformation, strain and stress distributions, and other data to be employed in the design process. An initial set of ply orientations and geometrical design of the prosthesis is inputted (step 220). A finite element analysis is carried out to determine design parameters such implane bending, torsional load, micromotion and deformations, normal stress distribution, and strain energy density (step 230). The results of these design parameters are compared with a set of target performance parameters (step 240). Depending on the comparisons, a determination is made to change the design of the composite prosthesis including the orientations of the plies (step 250) for repeating the step of finite element analysis (step 230). The iterative design process ends when the design parameters are with tolerance ranges of the target parameters (step 260). The fiber orientations of each layer and the combined stack sequence thus provide a very useful design parameter for a prosthesis designer to systematically determine the most fitting device. The details of the finite element analyses and the data comparison processes are summarized in two papers submitted to be published. These two papers are incorporated herein as references and enclosed with this continuation-in-part (CIP) Application as reference attachments. A prior art U.S. Pat. No. 5,064,439 (Chang et al.) by the same inventor of the present invention is also incorporated herein as a reference to provide additional background information.

☐ 11. Document ID: US 6258571 B1

L3: Entry 11 of 18 File: USPT Jul 10, 2001

US-PAT-NO: 6258571
DOCUMENT-IDENTIFIER: US 6258571 B1

TITLE: High throughput DNA sequencing vector

DATE-ISSUED: July 10, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Chumakov; Ilya	Vaulx-le-Penil			FR
Tanaka; Hiroaki	Antony			FR

US-CL-CURRENT: 435/91.42; 435/320.1, 435/473, 435/6, 435/91.1, 536/23.1, 536/23.2, 536/24.1, 536/24.3

ABSTRACT:

High throughput DNA sequencing vectors for generating nested deletions using enzymatic techniques and/or transposition-based techniques are disclosed. Methods of constructing contigs of long DNA sequences and methods of generating nested deletions are also disclosed. A truncated lacZ derivative useful in measuring the copy number of the lacZ derivative in a host cell is also disclosed.

26 Claims, 27 Drawing figures
Exemplary Claim Number: 1
Number of Drawing Sheets: 23

L3: Entry 11 of 18 File: USPT Jul 10, 2001

DOCUMENT-IDENTIFIER: US 6258571 B1
TITLE: High throughput DNA sequencing vector

Application Filing Year (1):
1999

Detailed Description Text (221):
The existence within the human genome of multiple repeats limits the number of clones with unique sequences at both ends and the efficiency of the OSS method. For instance, two distinct sequences can appear to overlap because they contain an ALU repeat. A practical way of avoiding this problem is to compare all primary produced sequences to a database of all known human repeats, and to mask all bases corresponding to the repeats. A masked base is declared to be not useful for the contigation. However, some of the primary sequences may contain a large part of a repeated sequence, preventing them from being contigated to the other sequences. Thus, when the repeat rate is high, a significant portion of the determined sequences can be lost for contigation. In such instances, the information necessary for the OSS method is lost and the gain of efficiency relative to the shotgun strategy is also lost. This phenomenon was observed empirically during practical application of the OSS strategy and in the computer simulation provided below.

☐ 12. Document ID: US 6228599 B1

L3: Entry 12 of 18 File: USPT May 8, 2001

US-PAT-NO: 6228599
DOCUMENT-IDENTIFIER: US 6228599 B1

TITLE: Antibody specific for homogalacturonan

DATE-ISSUED: May 8, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Knox; John Paul	Leeds			GB
Willats; William George Tycho	Ilklay			GB
Mikkelsen; Jorn Dalgaard	Hvidovre			DK

US-CL-CURRENT: 435/7.1; 435/810, 530/387.1, 530/387.3, 530/387.5

ABSTRACT:

Antibodies specific to pectin, specifically to homogalacuronin capable of recognizing a certain motif on the pectin structure were produced. These antibodies can be used alone or linked to a detectable moiety. They can be used in an assay or can be used to produce a food.

4 Claims, 6 Drawing figures
Exemplary Claim Number: 1
Number of Drawing Sheets: 4

L3: Entry 12 of 18 File: USPT May 8, 2001

DOCUMENT-IDENTIFIER: US 6228599 B1
TITLE: Antibody specific for homogalacturonan

Application Filing Year (1):
1999

Brief Summary Text (124):
FILTER--Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (1993) Computers and Chemistry 17:149-163, or segments consisting of short-periodicity internal repeats, as determined by the XNU program of Claverie & States (1993) Computers and Chemistry 17:191-201, or, for BLASTN, by the DUST program of Tatusov and Lipman (see Internet websit:www.ncbi.nlm.nih.gov). Filtering can eliminate statistically significant but biologically uninteresting reports from the blast output (e.g., hits against common acidic-, basic- or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.

Full | Title | Citation | Front | Review | Classification | Date | Reference | Claims | KWMC | Draw D

☐ 13. Document ID: US 6197932 B1

L3: Entry 13 of 18 File: USPT Mar 6, 2001

US-PAT-NO: 6197932
DOCUMENT-IDENTIFIER: US 6197932 B1

TITLE: Modulators of actin

DATE-ISSUED: March 6, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
King; Mary-Claire	Seattle	WA		
Lynch; Eric D.	Seattle	WA		
Lee; Ming K.	Bothell	WA		
Morrow; Jan E.	Seattle	WA		
Welcsh; Piri L.	Seattle	WA		
Leon; Pedro E.	San Jose			CR

US-CL-CURRENT: 530/350; 424/185.1, 436/86

ABSTRACT:

The invention provides methods and compositions which find use, inter alia, for modulating the stabilization of actin filaments. The compositions may comprise one or more polypeptide moieties derived from a novel human diaphanous polypeptide and/or one or more nucleic acid moieties derived from a novel human diaphanous gene or gene transcript. The invention also provides agents which specifically modify

the binding of a natural human diaphanous gene or gene product with a natural binding target thereof, isolated human diaphanous hybridization probes and primers capable of specifically hybridizing with the disclosed human diaphanous genes, human diaphanous-specific binding agents such as specific antibodies, and methods of making and using the subject compositions in diagnosis, therapy and in the biopharmaceutical industry.

19 Claims, 0 Drawing figures
Exemplary Claim Number: 1
L3: Entry 13 of 18 File: USPT Mar 6, 2001

DOCUMENT-IDENTIFIER: US 6197932 B1
TITLE: Modulators of actin

Application Filing Year (1):
1999

Detailed Description Text (20):
6. The SeqHelp program incorporates several sequence analysis programs and creates output in HTML files for browsing with any WWW browser (Lee et al Genomics submitted). The core programs used by Seqhelp are PHRED to read the ABI sequence files and assign bases, PHRAP to generate contigs of overlapping sequences, Repeat Masker (Arian Smit) to identify and mask common repetative elements prior to database searching, and BLAST (Altschul S, Gish W, Miller W, Myers E, Lipman D J Mol Biol 215:403-410 (1990)) comparison of project specific sequences to the NR and dbEST databases at the NCBI. An example of the SeqHelp output for analysis of the BRCA1 genomic region is available online at <hyper text transfer protocol://polaris.mbt.washington.edu>

Full | Title | Citation | Front | Review | Classification | Date | Reference | Claims | KWMC | Draw De

☐ 14. Document ID: US 6092070 A

L3: Entry 14 of 18 File: USPT Jul 18, 2000

US-PAT-NO: 6092070
DOCUMENT-IDENTIFIER: US 6092070 A

TITLE: Method and system for lossless date compression and fast recursive expansion

DATE-ISSUED: July 18, 2000

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Belcea; John Martin	Morristown	NJ		

US-CL-CURRENT: 707/101; 341/50

ABSTRACT:

A highly effective method for operating data processing equipment to achieve data compression with high coding and storage efficiency and a method and apparatus for

fast data retrieval while preserving full information content of the source data. This compressing method was used to successfully reduce the U.S. Geological Survey Database from 9.4 gigabytes to 800 megabytes, a reduction of over 90%. The compression method is an iterative and recursive process. At each iteration a data element is read into a buffer and then the pair formed by the last two elements in the buffer is checked against the rest of buffer. If a match is found in the buffer, the second element of the data element pair is removed and the first element is replaced by an index that indicates the sequential location in the buffer when the matching pair is found. The search for a matching pair is then repeated using the last two elements now in the buffer. When a matching pair is not found a new data element is added to the buffer and the whole process is repeated. After the last data element is entered in the buffer, the buffer is copied to an output file where the data elements are stored as is, and the location index is stored using fewer bits.

9 Claims, 7 Drawing figures
Exemplary Claim Number: 1
Number of Drawing Sheets: 5

L3: Entry 14 of 18 File: USPT Jul 18, 2000

DOCUMENT-IDENTIFIER: US 6092070 A
TITLE: Method and system for lossless date compression and fast recursive expansion

Application Filing Year (1):
1996

Brief Summary Text (13):
My invention affords a highly effective method for data compression to achieve high coding and storage efficiency and as systems and method for fast data retrieval while preserving full information content of the source data. My method has successfully reduced the U.S. Geological Survey Database from 9.4 gigabytes to 800 megabytes, a reduction of over 90%. The compression method of my invention is an iterative process. At each iteration a data element is read into a buffer and then the pair formed by the last two elements in the buffer is checked against the rest of buffer. If a match is found in the buffer, the second element of the data element pair is removed and the first element is replaced by an index that indicates the location in sequence in the buffer where the matching pair is found. The search for a matching pair is then repeated using the last two elements now in the buffer. When a matching pair is not found, a new data element is added to the buffer and the whole process is repeated. After the last data element is entered in the buffer, the buffer is copied to an output file. Data elements are stored using only the number of bits necessary to represent the data elements and the location index is stored using the fewest bits necessary to represent the location index number.

☐ 15. Document ID: US 6022716 A

L3: Entry 15 of 18 File: USPT Feb 8, 2000

US-PAT-NO: 6022716
DOCUMENT-IDENTIFIER: US 6022716 A
**** See image for Certificate of Correction ****

TITLE: High throughput DNA sequencing vector

DATE-ISSUED: February 8, 2000

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Chumakov; Ilya	Vaulx-le-Penil			FR
Tanaka; Hiroaki	Antony			FR

US-CL-CURRENT: 435/91.42; 435/320.1, 435/476, 435/91.4, 435/91.5, 536/23.2, 536/23.7, 536/24.1, 536/24.33

ABSTRACT:

High throughput DNA sequencing vectors for generating nested deletions using enzymatic techniques and/or transposition-based techniques are disclosed. Methods of constructing contigs of long DNA sequences and methods of generating nested deletions are also disclosed. A truncated lacZ derivative useful in measuring the copy number of the lacZ derivative in a host cell is also disclosed.

107 Claims, 19 Drawing figures
Exemplary Claim Number: 1
Number of Drawing Sheets: 23

L3: Entry 15 of 18 File: USPT Feb 8, 2000

DOCUMENT-IDENTIFIER: US 6022716 A
**** See image for Certificate of Correction ****
TITLE: High throughput DNA sequencing vector

Application Filing Year (1):
1998

Detailed Description Text (225):
The existence within the human genome of multiple repeats limits the number of clones with unique sequences at both ends and the efficiency of the OSS method. For instance, two distinct sequences can appear to overlap because they contain an ALU repeat. A practical way of avoiding this problem is to compare all primary produced sequences to a database of all known human repeats, and to mask all bases corresponding to the repeats. A masked base is declared to be not useful for the contigation. However, some of the primary sequences may contain a large part of a repeated sequence, preventing them from being contigated to the other sequences. Thus, when the repeat rate is high, a significant portion of the determined sequences can be lost for contigation. In such instances, the information necessary for the OSS method is lost and the gain of efficiency relative to the shotgun strategy is also lost. This phenomenon was observed empirically during practical application of the OSS strategy and in the computer simulation provided below.

Full | Title | Citation | Front | Review | Classification | Date | Reference | Claims | KWIC | Draw D

☐ 16. Document ID: US 5985574 A

L3: Entry 16 of 18 File: USPT Nov 16, 1999

US-PAT-NO: 5985574
DOCUMENT-IDENTIFIER: US 5985574 A

TITLE: Modulators of actin

DATE-ISSUED: November 16, 1999

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
King; Mary-Claire	Seattle	WA		
Lynch; Eric D.	Seattle	WA		
Lee; Ming K.	Bothall	WA		
Morrow; Jan E.	Seattle	WA		
Welcsh; Piri L.	Seattle	WA		
Leon; Pedro E.	San Jose			CR

US-CL-CURRENT: 435/6; 435/252.3, 435/320.1, 435/325, 435/69.1, 435/86, 530/387.1

ABSTRACT:

The invention provides methods and compositions which find use, inter alia, for modulating the stabilization of actin filaments. The compositions may comprise one or more polypeptide moieties derived from a novel human diaphanous polypeptide and/or one or more nucleic acid moieties derived from a novel human diaphanous gene or gene transcript. The invention also provides agents which specifically modify the binding of a natural human diaphanous gene or gene product with a natural binding target thereof, isolated human diaphanous hybridization probes and primers capable of specifically hybridizing with the disclosed human diaphanous genes, human diaphanous-specific binding agents such as specific antibodies, and methods of making and using the subject compositions in diagnosis, therapy and in the biopharmaceutical industry.

6 Claims, 0 Drawing figures
Exemplary Claim Number: 1

L3: Entry 16 of 18 File: USPT Nov 16, 1999

DOCUMENT-IDENTIFIER: US 5985574 A
TITLE: Modulators of actin

Application Filing Year (1):
1998

Detailed Description Text (20):
6. The SeqHelp program incorporates several sequence analysis programs and creates output in HTML files for browsing with any WWW browser (Lee et al Genomics submitted). The core programs used by Seqhelp are PHRED to read the ABI sequence files and assign bases, PHRAP to generate contigs of overlapping sequences, Repeat Masker (Arian Smit) to identify and mask common repetative elements prior to database searching, and BLAST (Altschul S, Gish W, Miller W, Myers E, Lipman D J Mol Biol 215:403-410 (1990)) comparison of project specific sequences to the NR and dbEST databases at the NCBI. An example of the SeqHelp output for analysis of the BRCA1 genomic region is available online at <http://polaris.mbt.washington.edu>7.

☐ 17. Document ID: US 5940825 A

L3: Entry 17 of 18 File: USPT Aug 17, 1999

US-PAT-NO: 5940825
DOCUMENT-IDENTIFIER: US 5940825 A

TITLE: Adaptive similarity searching in sequence databases

DATE-ISSUED: August 17, 1999

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Castelli; Vittorio	White Plains	NY		
Li; Chung-Sheng	Ossining	NY		
Yu; Philip Shi-lung	Chappaqua	NY		

US-CL-CURRENT: 707/6; 707/2, 707/3, 707/7

ABSTRACT:

A computer system and method for performing similarity searches which is phase and scale insensitive and which allows similarity searches to be performed at a semantic level. Each sequence in a database is preferably segmented at multiple projections and/or resolution levels. The sequences may represent object having multi-dimensional features such as temporal and/or spatial-temporal data. Preferably, the segmenting logic starts with the finest resolution, and each sequence is parsed into a number of disjointed segments, wherein each segment has uniform features. The uniform features could be segments having a constant slope, or waveform segments representable by a single function. The segments may then be re-sampled into a fixed length vector with appropriate normalization. A label may also be assigned to each segment via conventional clustering/classification methods. The above steps are iterated at successive projections and/or resolution levels until each sequence in the database has been independently segmented and clustered. Thus, the labels are preferably extracted in a pseudo-hierarchical manner in which the label of the lowest resolution representation of the sequence is extracted first. The representation of each time series at various resolutions and/or projections captures different characteristics of the same time series (or 2D/3D objects). Recall that each segment represents a region having uniform features. The segmentation at each individual resolution and/or projection thus enables recognition or emphasis of different characteristics within segments having uniform features.

39 Claims, 9 Drawing figures
Exemplary Claim Number: 15
Number of Drawing Sheets: 9

L3: Entry 17 of 18 File: USPT Aug 17, 1999

DOCUMENT-IDENTIFIER: US 5940825 A
TITLE: Adaptive similarity searching in sequence databases

Application Filing Year (1):
1996

Detailed Description Text (12):
FIG. 6 is a flowchart of an alternative iterative refinement method for the database segmenting and clustering logic of FIG. 4. As depicted, the variable COST1 may be initially assigned to a very high or even infinite value. In step 601, a seed segmentation can be used to parse the entire database. Seed segmentation is well known in the art. See for example, Shatkay and Zdonik, "Approximate Queries and Representations for Large Data Sequences," Proc. ICDE, pp. 536-545, February 1996. In step 602, each segment is re-sampled, and in step 603, clustered (step 403 in FIG. 4). In step 604, the performance of each cluster configuration is evaluated based on a specific performance metric. The performance metric of a cluster can be defined, for example, as the mean variance of the clusters. In this case, the clustering variance decreases as the performance of the configuration improves. In step 605, the difference between the performance metric of the current configuration and the previous configuration is calculated. In step 606, if the difference is smaller than a certain predefined threshold, this cluster configuration is accepted as the final output cluster. Otherwise, in step 607, the cost (performance metric) of the current configuration is reassigned to the cost of the previous configuration. In step 608, a perturbation of the initial segmentation is generated to obtain new clustering results. This perturbation is accepted if the perturbation improves the clustering results. This process repeats until the clustering performance levels off, in step 606.

☐ 18. Document ID: US 5802208 A

L3: Entry 18 of 18 File: USPT Sep 1, 1998

US-PAT-NO: 5802208
DOCUMENT-IDENTIFIER: US 5802208 A

TITLE: Face recognition using DCT-based feature vectors

DATE-ISSUED: September 1, 1998

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Podilchuk; Christine Irene	Bridgewater	NJ		
Zhang; Xiaoyu	Piscataway	NJ		

US-CL-CURRENT: 382/224; 382/115, 382/250, 382/253

ABSTRACT:

A method and apparatus for recognition of objects such as faces in images using signal compression techniques (i.e., coding techniques) in which a portion of the image which includes the object to be recognized (e.g., the face) is coded, and the resultant coded data is matched against previously coded and stored training data which makes up a known object database. A given object in an input image signal is matched to one of a plurality of known objects stored in a database, wherein the

stored representation of each of the known objects comprises a codebook generated based on training image signals comprising the known object. A first illustrative embodiment comprises the steps of decomposing the given object into blocks; performing a plurality of encodings of the given object, each encoding comprising coding the object with use of one of the codebooks; determining a coding error for each encoding; and matching the given object to one of the known objects based on the coding errors. A second illustrative embodiment comprises the steps of decomposing the given object into blocks; generating a codebook corresponding to the given object based on the blocks; comparing the codebook corresponding to the given object with the codebooks corresponding to each of the known objects; and matching the given object to one of the known objects based on the comparison of the codebooks.

18 Claims, 5 Drawing figures
Exemplary Claim Number: 1
Number of Drawing Sheets: 2

L3: Entry 18 of 18 File: USPT Sep 1, 1998

DOCUMENT-IDENTIFIER: US 5802208 A
TITLE: Face recognition using DCT-based feature vectors

Application Filing Year (1):
1996

Detailed Description Text (8):
The illustrative procedure shown in FIG. 3 comprises an iterative process that generates a codebook by repeatedly matching the training vectors to a sequence of "intermediate" codebooks, C.sup.p.sub.0, C.sup.p.sub.1, . . . , C.sup.p.sub.m, . . . , modifying the codebook on each iteration (i.e., replacing codebook C.sup.p.sub.m with improved codebook C.sup.p.sub.m+1), until a terminating criterion is met. The codebook which results from the final iteration then advantageously becomes the codebook which is used to represent the particular individual's face in the database.

Term	Documents
@AY	23272276
(2 AND (@AY < "2000")).PGPB,USPT,EPAB,JPAB,DWPI.	18
(L2 AND @AY<2000).PGPB,USPT,EPAB,JPAB,DWPI.	18